STIC Search Report
Biotech-Chem Library

USPTO

TO: William W Moore
Location: CM1/10A09/10D01
Art Unit: 1652
 Tuesday, July 01, 2003

Case Serial Number: 014896

From: Toby Port
Location: Biotech-Chem Library
CM1-6A04
Phone: 308-3534

toby.port@uspto.gov

Search Notes

Dear Examiner Moore,

Here are the results of your search.
Please feel free to contact me if you have any questions.

Toby Port

STIC SEARCH RESULTS FEEDBACK FORM

Questions about the scope or the results of the search? Contact **the searcher or contact**:

Mary Hale, Information Branch Supervisor
308-4258, CM1-1E01

Voluntary Results Feedback Form

➢ *I am an examiner in Workgroup:* [   ]   *Example: 1610*

➢ *Relevant prior art **found**, search results used as follows:*

☐ 102 rejection

☐ 103 rejection

☐ Cited as being of interest.

☐ Helped examiner better understand the invention.

☐ Helped examiner better understand the state of the art in their technology.

   Types of relevant prior art found:

   ☐ Foreign Patent(s)

   ☐ Non-Patent Literature
   (journal articles, conference proceedings, new product announcements etc.)

➢ *Relevant prior art **not found:***

   ☐ Results verified the lack of relevant prior art (helped determine patentability).

   ☐ Results were not useful in determining patentability or understanding the invention.

Comments:

Drop off or send completed forms to STIC/Biotech-Chem Library CM1 – Circ. Desk

sira
Search and Information
Resources Administration

97477

From:	Moore, William
Sent:	Wednesday, June 25, 2003 3:31 PM
To:	STIC-Biotech/ChemLib
Subject:	Sequence Search request for 10/014,896

Please search SEQ ID NO:1 in 10/014,896 in the following ways:

     1. N2N standard, Search SEQ ID NO:1 in the EST, GenEmbl, N-GeneSeq, US patented, and USPGpublished nucleic acid sequences.

     2. N2N interference, Search SEQ ID NO:1 in the US pending main and pending new application nucleic acid sequences.

     3. N2P standard, Search SEQ ID NO:1 **against** the PIR, SwissProt, SPTREMBL, A-GeneSeq, US patented, and USPGpublished amino acid sequences.

     4. N2P interference, Search SEQ ID NO:1 **against** the US pending main and pending new application amino acid sequences.

Thanks,

William W. Moore, Art Unit 1652

Telephone: 703.308.0583

Office: 10A09 Crystal Mall 1

Mailbox: 10D01 Crystal Mall 1

william.moore@uspto.gov

Pending Nucleic Acid and/or Pending Amino Acid database searches now generate two sets of results. These databases were split into to two parts to reduce the time needed to update the databases daily. The split freed up more machine time for processing searches.

Searches run against the Nucleic Acid Pending database produce two sets of results, with the extensions, **.rnpm** and **.rnpn**
Searches run against the Amino Acid Pending database produce two sets of results, with the extensions, **.rapm** and **.rapn**

The Pending database search results should not be left in the case because they contain data that is confidential.

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 27, 2003, 10:58:03 ; Search time 357.5 Seconds
(without alignments)
5442.810 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 2762
Sequence: 1 atggctcaggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 4569144 seqs, 644733110 residues

Total number of hits satisfying chosen parameters: 9138288

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-O=/cgn2_1/USPTO_spool/US10014896//runat_27062003_104415_10437/app_query.fasta_1.1671
-DB=Pending_Patents_AA_Main -QFMT=fastan -SUFFIX=n2p.rapm -MINMATCH-0.1
-LOOPCT-0 -LOOPEXT=0 -UNITS=bits -START-1 -END=-1 -MATRIX-blosum62
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN-200000000000 -USER=US10014896_@CGN_1_562_@runat_27062003_104415_10437
-NCPU-6 -ICPU-3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK-100
-LONGLOG - DEV_TIMEOUT=120 -WARN_TIMEOUT-30 -THREADS=1 -XGAPOP=10 -XGAPEXT-0.5
-FGAPOP=6 -FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP=6 -DELEXT=7

Database :
Pending_Patents_AA_Main:*
1: /cgn2_6/ptodata/1/paa/pcrus_COMB.pep:*
2: /cgn2_6/ptodata/1/paa/US06_COMB.pep:*
3: /cgn2_6/ptodata/1/paa/US07_COMB.pep:*
4: /cgn2_6/ptodata/1/paa/US080_COMB.pep:*
5: /cgn2_6/ptodata/1/paa/US081_COMB.pep:*
6: /cgn2_6/ptodata/1/paa/US082_COMB.pep:*
7: /cgn2_6/ptodata/1/paa/US083_COMB.pep:*
8: /cgn2_6/ptodata/1/paa/US084_COMB.pep:*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep:*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep:*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep:*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep:*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep:*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep:*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep:*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep:*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep:*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep:*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep:*
20: /cgn2_6/ptodata/1/paa/US096_COMB.pep:*
21: /cgn2_6/ptodata/1/paa/US097_COMB.pep:*
22: /cgn2_6/ptodata/1/paa/US098_COMB.pep:*
23: /cgn2_6/ptodata/1/paa/US099_COMB.pep:*
24: /cgn2_6/ptodata/1/paa/US100_COMB.pep:*
25: /cgn2_6/ptodata/1/paa/US101_COMB.pep:*
26: /cgn2_6/ptodata/1/paa/US102_COMB.pep:*
27: /cgn2_6/ptodata/1/paa/US60_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2546	92.2	502	24	US-10-014-896-2	Sequence 2, Appl
2	2538	91.9	502	27	US-60-345-785-2	Sequence 2, Appl
3	1824	66.0	361	24	US-10-014-896-4	Sequence 4, Appli
4	1078	39.0	236	27	US-60-213-845-350	Sequence 350, App
5	796.5	28.8	408	23	US-09-902-540-10715	Sequence 10715, A
6	794	28.7	154	1	US-60-311-261-2481	Sequence 2481, Ap
7	562.5	20.4	120	1	PCT-US02-09921-1008	Sequence 1008, Ap
8	562.5	20.4	120	27	US-60-311-261-2481	Sequence 2481, Ap
9	535	19.4	581	25	US-10-179-131-8920	Sequence 8920, Ap
10	532	19.3	616	16	US-09-248-796-17521	Sequence 17521, A
11	532	19.3	616	27	US-60-096-409-17521	Sequence 17521, A
12	504.5	18.3	623	25	US-10-179-131-7140	Sequence 7140, Ap
13	451	16.3	93	18	US-08-904-468-192	Sequence 192, App
14	451	16.3	93	19	US-09-547-599C-2910	Sequence 2910, Ap
15	410	14.8	79	27	US-60-195-044-179	Sequence 179, App
16	361.5	13.1	443	21	US-09-791-537-28973	Sequence 28973, A
17	334	12.1	468	24	US-10-015-127-11248	Sequence 11248, A
18	317.5	11.5	445	24	US-09-791-537-12737	Sequence 12737, A
19	309.5	11.2	448	21	US-09-791-537-96235	Sequence 96235, A
20	307	11.1	359	18	US-09-417-507-22502	Sequence 22502, A
21	302.5	11.0	228	16	US-09-248-796-17522	Sequence 17522, A
22	302.5	11.0	228	27	US-60-096-409-17522	Sequence 17522, A
23	297	10.8	112	1	PCT-US02-25485-1140	Sequence 1140,
24	297	10.8	112	27	US-60-311-261-1141	Sequence 1141, Ap
25	275.5	10.0	441	20	US-09-738-626-6613	Sequence 6613, Ap
26	274	9.9	422	20	US-09-602-777A-202	Sequence 202, App
27	274	9.9	422	20	US-09-606-740A-372	Sequence 372, App
28	253.5	9.2	401	20	US-09-614-150-14709	Sequence 14709, A
29	253.5	9.2	401	21	US-09-791-537-12876	Sequence 12876, A
30	253.5	9.2	401	27	US-60-173-464-12029	Sequence 12029, A
31	253.5	9.2	401	14	US-60-191-637-14750	Sequence 14750, A
32	253.5	9.2	814	27	US-60-191-681-11627	Sequence 11627, A
33	233.5	8.5	233	14	US-60-167-217-14782	Sequence 14782, A
34	228	8.3	400	20	US-09-614-150-25302	Sequence 25302, A
35	228	8.3	400	21	US-09-791-537-28760	Sequence 28760,
36	228	8.3	400	27	US-09-791-537-128760	Sequence 128760, A
37	228	8.3	400	20	US-60-191-637-25418	Sequence 25418, A
38	228	8.3	400	27	US-60-191-681-20027	Sequence 20027, A
39	224.5	8.1	408	21	US-09-614-150-25308	Sequence 25308, A
40	224.5	8.1	408	20	US-09-791-537-128759	Sequence 128759, A
41	224.5	8.1	408	27	US-60-173-464-20944	Sequence 20944, A
42	224.5	8.1	408	27	US-60-191-637-25424	Sequence 25424, A
43	224.5	8.1	408	27	US-60-191-681-20033	Sequence 20033, A
44	220	8.0	359	27	US-60-167-217-14769	Sequence 14769, A
45	220	8.0	359	27	US-60-173-464-12011	Sequence 12011, A

RESULT 1
US-10-014-896-2
Sequence 2, Application US/10014896
GENERAL INFORMATION:
APPLICANT: Yu, Xuanchuan
APPLICANT: Miranda, Maricar
APPLICANT: Friddle, Carl Johan
TITLE OF INVENTION: Novel Human Proteases and
TITLE OF INVENTION: Polynucleotides Encoding the Same
FILE REFERENCE: LEX-0280-USA
CURRENT APPLICATION NUMBER: US/10/014,896
CURRENT FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 60/255,567
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 4

ALIGNMENTS

(handwritten annotation in margin: "Hit No. 2 is too late; too No. 2 is too dis that" — illegible)

```
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-014-896-1
US-10-014-896-2

Alignment Scores:
Pred. No.:
Score:                     6.71e-231
Matches:                   2546.00              Length:        502
Percent Similarity:        100.00%              Matches:       502
Best Local Similarity:     100.00%              Conservative:  0
Query Match:               92.18%               Mismatches:    0
DB:                        24                   Indels:        0
                                                Gaps:          0
```

US-10-014-896-1 (1-1509) x US-10-014-896-2 (1-502)

```
Qy   1 ATGGCTCAGCCACCGGTTGCGGTTTGCGGTGGTCGCTTCCTTAGTTTTCCCT 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaGlnArgCysValCysValLeuAlaLeuValAlaMetLeuLeuLeuValPhePro 20

Qy  61 ACCGTCTCCAGATCGANTGGGCCCGAGGAGCGGGGAGCATCAAAGGGCGTCGGAATCCCT 120
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  21 ThrValSerArgSerMetGlyProArgSerGlyLeuHisGlnArgAlaSerArgIlePro 40

Qy 121 TCTCACTTCACGAAAGAGAAGAACGCGTCGGATGAAAGAGGCGGCGAAAGGTGCCATCCAG 180
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  41 SerGlnPheSerLysGluGlyAlaAlaMetLysGluAlaAlaLysGlyAlaIleGln 60

Qy 181 ATTCCAACAGTGACTTTAGCTCGAGAGTCCAATACTACAGCCCTGGCCTGAGTTCGGA 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly 80

Qy 241 AAAATACATTCAAAAAGTCTTACACAGTGGTCAGCAGCTTTATCCAGCATGAAGTC 300
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  81 LysTyrIleHisLysValPheProThrValValSerThrSerPheIleGlnHisGluVal 100

Qy 301 GTGGAAGAGTATAGCCACCGTTCACTATCCCAAGGCTCGGACCCCCAGCTTGCAGCCCTAC 360
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspSerProSerLeuGlnProTyr 120

Qy 361 CTGCTGATGGCCTCACTTTGATGTGGTGGCTCTGCCCTGAACAAGGCTGGGAGGTGCCCCCA 420
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro 140

Qy 421 TTCTCTGGGGTTGGGAGCGGTGATGGGCCTGCTCATCTATGTGTCGGGGCACACTGACGACGAAGAAC 480
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141 PheSerGlyLeuGlyArgGlyAspGlyValIleTyrGlyArgGlyThrLeuAspAspLysAsn 160

Qy 481 TCTGTGATGGCATTACTGCCAGGCCTTGGAGACTCCCGCTGCGGAAGTACATCCCCGA 540
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161 ServalMetAlaLeuLeuGlnAlaLeuGluLeuteuLeuArgLysTyrIleProArg 180

Qy 541 AGATCTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCAGGGACACGGGGCTCAGAGG 600
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ArgSerPheTheIleSerLeuGlyHisAspGluSerSerGlyTHrGlyAlaGlnArg 200

Qy 601 ATCTCAGCCCTGCTACAGTCAAAGGGCGTCCAGCTTAGCCTTCATTGTGGACGAGGGGGGC 660
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 201 IleSerAlaLeuLeuGlnSerArgGlyValGlnLeuAlaPheIleValAspGluGlyGly 220

Qy 661 TTCATCTTGGANGANATTTCATTCCTAACTTCAAGAAGCCCATCGCCTTGAATTGCAGTCTCTCA 720
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 221 PheIleLeuAspAspPheIleProAsnPheLysLysProIleAlaLeuAsnCysSerLeu 240

Qy 721 GAGAAGGGGTTCCATGGANGANATCCATGCTGCAAGTAAACATGACTTCAGGACCACTCTTCAGCT 780
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla 260

Qy 781 CCTCCAAAGGAGACAAGACAATTGGCATCCTTGCACGTCGTCGGATTGGAGCGAGACA 840
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 261 ProProLysGluThrSerIleGlyIleLeuAlaArgArgArgIleGlyAlaGlnThr 280
```

RESULT 2
US-60-345-785-2
; Sequence 2, Application US/6034578
; GENERAL INFORMATION:
; APPLICANT: CURTIS, Rory A.J.
; TITLE OF INVENTION: 58812, A Human Metallopeptidase Family
; TITLE OF INVENTION: Member and Uses Therefor
; FILE REFERENCE: MPI2002-003P1(M)
; CURRENT APPLICATION NUMBER: US/60/345,785
; CURRENT FILING DATE: 2002-01-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 502
; TYPE: PRT
; ORGANISM: Homo sapien
US-60-345-785-2

```
Alignment Scores:
Pred. No.:
Score:                     3.83e-230
Matches:                   2538.00              Length:        502
Percent Similarity:        99.80%               Matches:       501
Best Local Similarity:     99.80%               Conservative:  0
Query Match:               91.89%               Mismatches:    1
DB:                        27                   Indels:        0
                                                Gaps:          0
```

```
Qy  841 CCAATGCCTATCATATATTTGGAAGCGGGGACAGTGTGTGACTGTATTGCAGCAACTGGCAAAT 900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  281 ProMetProIleIlePheGlySerGlyThrValValThrValLeuGlnLeuGlnAlaAsn 300

Qy  901 GAGTTTCCCTTCCCGTCAAATAATAAATCCTGAGCAACCCARGGCTATTGAACACACTAATA 960
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GluPheProPheProValAsnIleIleLeuSerAsnProArgLeuPheGlyLeuPhe 320

Qy  961 AGCAGGTTTAGGGAGGAGAAATCCTTAACCAAATGCAAATAATCAGGACGACCACGGCGACTC 1020
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  321 SerArgPheMetGluArgArgAsnProLeuThrAsnAlaIleIleArgThrThrThrAlaLeu 340

Qy 1021 ACCATATTCAAAGCACGGGGTCAAGTTCAATGTCATCCCCCAGTTGGCCCAGGCCCACGTC 1080
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  341 ThrIlePheLysAlaGlyValLysPheAsnValIleProGlnProValAlaGlnAlaThrVal 360

Qy 1081 AACTTCCGGATTCACCCTGGACAGACAGCAGTCCAAGAGGGTCCTAGAAGACTCACGAAGAAACATT 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AsnPheArgIleHisProGlyGlnThrValGlnGluValLeuGluThrLysAsnIle 380

Qy 1141 GTGGGCTGATAACAGAGTGCCAGTTCCAATGTGTTGGAGTGGCCTTTGACCCCTCCCCGTAGC 1200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  381 ValAlaAlaAspAsnArgValGlnPheHisIleValLeuSerAlaPheAspProLeuProValSer 400

Qy 1201 CCTTCTGATGACAAGGCCTTGGGCTACCAGCTGCTCCCGCCAGACCGTACAGTCGTCTTC 1260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  401 ProSerAspSerLysAlaLeuGlyTyrGlnLeuLeuArgGlnAspArgThrValValPhe 420

Qy 1261 CCGGAAGTCAATATTATTACTGCCCCCAGTTACTTCTATTGGGAACAAGGCAGCCCCATTCTTT 1320
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 ProGluValAsnIleIleThrAlaProValThrSerIleGlyAsnThrHisAspSerArgPhe 440

Qy 1321 ACAAAACCTCACCACTCGCACNTCTACAGGTTCTACCCCATCTACAATACAGCCTGAAGACTTC 1380
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  441 ThrAsnLeuThrThrGlyIleTyrArgPheTyrProIleTyrAsnThrAlaGlnProGluAspPhe 460

Qy 1381 AAACGCATCCATGGAGGTCAAGGAGAAAATCTCAGTCGCCAAGCCTANGAGAGACCCAAGTGAA 1440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  461 LysArgIleHisGlyValAsnGluLysIleSerValAlaLysAlaTyrGluThrGlnValLys 480

Qy 1441 TTCATCTTTGAGTTGATTCAGAATGCTGACACGAAGCAGGAGCAGTTTTCTCACCTTGCAC 1500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 PheIlePheGluLeuIleGlnAsnAlaAspThrLysGlnGluGlnPheSerHisLeuHis 500

Qy 1501 AAACTG 1506
        ||||||
Db  501 LysLeu 502
```

US-10-014-896-1 (1-1509) x US-60-345-785-2 (1-502)

```
Qy      1  ATGGCTCAGCGGTGCGTTTGCGTGCTGGCCTATGCTGCTCCTAGTTTTCCT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MetAlaGlnArgCysValCysValLeuAlaLeuAlaLeuLeuValPhePro  20

Qy     61  ACCGGTCTCCAGATCGATGGGCCCGAGGAGGCGGGGCATCAAAGGCGGTCCCGGAATCCCT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  ThrValSerArgSerMetGlyProArgSerGlyValHisGlnArgAlaSerArgIlePro  40

Qy    121  TCTCAGTTCAGCAAAGAGAACCGGTCCCGATCAAAAGAGGTTGCCATCCAG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41  SerGlnPheSerLysGluGluArgValAlaMetLysGluValGlyAlaIleGln  60

Qy    181  ATTCCAACACTGACTTTTAGCTCTGAGAAGTCCAAATACTACAGCCCTGGCTGGAGTTCGGA  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluProGly  80

Qy    241  AAATACATTCATAAAGTCTTTCCTACAGTGGTCGAGCACCAGTTTATCACGGCGATGAAGTC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81  LysTyrIleHisLysValPheProThrValValGluHisGlnPheIleGlnHisHisGlu  100

Qy    301  GTGGAAGAGTATAGGCCACCTGTTCACTATCCCAAGGCTGGACCCCAGCTTTGCAGCCCTAC  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101  ValGlnGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuAlaProTyr  120

Qy    361  CTGCTGATGGCTCACTTTGATGTGGTGGCCTGCCCCTCGAAGAAGGCTGGTGCCCCCA  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  LeuLeuMetAlaHisPheAspValValAlaProLeuArgProGluGluGlyTrpValProPro  140

Qy    421  TTTCTCTGGGTTTGGACCGTGATGGGCCGTCACTATCGGGTGCACACTGACAGCAGAAC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141  PheSerGlyLeuGlyArgAspGlyValValIleTyrGlyValHisThrLeuAspAspLysAsn  160

Qy    481  TCTGTGGATGGCATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAGTACATCCCCGA  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161  SerValMetAlaLeuLeuGlnAlaLeuGluLeuLeuLeuIleArgGlyTyrIleProArg  180

Qy    541  AGATCTTTCTTCATTTCTCTGGGCCATGANTGAGGAGTCATCAGGGACAGAGGGGGTCAGAGG  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  ArgSerPhePheIleSerLeuGlyHisAspGluSerSerGlyThrGlyAlaGlnArg  200

Qy    601  ATCTTCAGGCCCTGCTACAGTCAAGAGGGGGTCCAGCTAGCCTTCATTGTGGGACAGAGGGGC  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201  IleSerArgLeuLeuGlnSerArgGlyValGlnLeuAlaPheIleValAspGlyLuGlyGly  220

Qy    661  TTCATCTTGGATGGATTTCATTCCTAACTTCAAGAAGCCCATCGGCCTTGATTGCAGTCTCA  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221  PheIleLeuAspAspPheIleProAsnPheLysLysProIleAlaLeuIleAlaValSer  240

Qy    721  GAGAAGGGTCCATGAACCTCATGGTCTGCAAGTAAACATGACTTCAGGCCGCCACTCTTCAGCT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla  260

Qy    781  CCTCCAAAGAGGAGACAAGCAATTGGCACCCTTGCACGTGCTGCTGTCACCGGATTGGAGCAGACA  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261  ProProLysGluGluThrSerIleGlyIleLeuAlaLeuAlaAlaValSerArgGlueGlnThr  280

Qy    841  CCAATGCCTATCATATTTGAAGCCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAAT  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281  ProMetProIleIlePheGluAlaGlyThrValValThrValLeuGlnGlnLeuAlaAsn  300

Qy    901  GAGTTCCCTGCCTGTCTCAATATAAATCCTGAGCAACCCATAGGCTGTATTTGAACCACTTATA  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  GluPheProAlaValSerAsnIleIleLeuSerAsnProIleProGlyCysIleProLeuIle  320

Qy    961  AGCAGGGTTATTGAGAGAAAATCCTTAACCAATGCAATAATCAGGACCACCAGCGGCACTC  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    321  SerArgPheMetGluArgAsnProLeuThrAsnAlaIleIleArgThrThrThrThrAlaLeu  340

Qy   1021  ACCAATATTCAAAGCAGGGGTCCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCCCAGGTC  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 3
US-10-014-896-4
Sequence 4, Application US/10014896

```
; GENERAL INFORMATION:
;    APPLICANT: Yu, Xuanchuan
;    APPLICANT: Miranda, Maricar
;    APPLICANT: Friddle, Carl Johan
;    TITLE OF INVENTION: Novel Human Proteases and
;    TITLE OF INVENTION: Polynucleotides Encoding the Same
;    FILE REFERENCE: LEX-0280-USA
;    CURRENT APPLICATION NUMBER: US/10/014,896
;    CURRENT FILING DATE: 2001-12-11
;    PRIOR APPLICATION NUMBER: US 60/255,567
;    PRIOR FILING DATE: 2000-12-14
;    NUMBER OF SEQ ID NOS: 4
;    SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;    LENGTH: 361
;    TYPE: PRT
;    ORGANISM: Homo sapiens
US-10-014-896-4
```

Alignment Scores:

Pred. No.:	1
Score:	1.01e-162
Percent Similarity:	1824.00
Best Local Similarity:	93.99%
Query Match:	93.99%
DB:	66.04%

Length:	361
Matches:	361
Conservative:	360
Mismatches:	0
Indels:	1
Gaps:	22
	1
	24

US-10-014-896-1 (1-1509) x US-10-014-896-4 (1-361)

```
Qy      1  ATGGCTCAGCGGTGCGTTTGCGTGCTGGCCTATGCTGCTCCTAGTTTTCCT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MetAlaGlnArgCysValCysValLeuAlaLeuAlaLeuLeuValPhePro  20

Qy     61  ACCGGTCTCCAGATCGATGGGCCCGAGGAGGCGGGGCATCAAAGGCGGTCCCGGAATCCCT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  ThrValSerArgSerMetGlyProArgSerGlyValHisGlnArgAlaSerArgIlePro  40
```

Db   341 ThrIlePheLysAlaGlyValLysPheAsnValIleProProValAlaGlnAlaThrVal 360
```
Qy   1081  AACTTCCGGATTCACCCTGGACGAACAGTCCAAGAGGTCCTAGAACTCAGAAGAAGACATT  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  AsnPheArgIleHisProGlyGlnThrValGlnGluValGlnLeuThrLysAsnIle  380

Qy   1141  GTGGCTGATAACAGAGTCCAGTTCCARGTGTTGAGTGGCCTTGACCCCCGTCAGC  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    381  ValAlaAspAsnArgValGlnPheHisValLeuSerAlaPheAspProLeuProValSer  400

Qy   1201  CCTTCTGATGACAGAAGGCCTTGGGCTACCAGCTGCTCCGCCAGACCGTACAGTCCGTCTTC  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    401  ProSerAspAspArgLysProLeuGlyTyrGlnLeuLeuArgGlnThrValGlnSerValPhe  420

Qy   1261  CCGGAAGTGCAATAATTACTGCCCCAGTTACTTCTATTGGCAACACAGACAGCCGATTCTTT  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  ProGluValAsnIleThrAlaProValThrSerIleGlyAsnThrAspSerArgPhePhe  440

Qy   1321  ACAAACCTCACCACTGCGCATTCTACAGTTCTCACCCCATCTACACAGCCTGAAGACTTC  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    441  ThrAsnLeuThrThrGlyIleTyrArgPheTyrProIleTyrThrArgArgThrIleAspPhe  460

Qy   1381  AAACGGACCATCCATGGAGTCAACCAGAGAAAATCCAGTGCCAAGCCTATGAGACCCAAGTGAA  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    461  LysArgIleHisGlyValAsnGlnIleSerSerValGlnLysIleGlnThrGlnValLys  480

Qy   1441  TTCATCTTTGAGTTGATTCAGAAATGCTGACACGACAGGACCACAGACAGTTTCCACCTGCAC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  PheIlePheGluLeuLeuIleGlnAsnAlaAspThrAspGlnProValSerHisLeuHis  500

Qy   1501  AAACTG  1506
           ||||||
Db    501  LysLeu  502
```

```
Qy      1  ATGGCTCAGCGGTGCGTTTGCGTGCTGGCCTATGCTGCTCCTAGTTTTCCT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  MetAlaGlnArgCysValCysValLeuAlaLeuAlaLeuLeuValPhePro  20

Qy     61  ACCGGTCTCCAGATCGATGGGCCCGAGGAGGCGGGGCATCAAAGGCGGTCCCGGAATCCCT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21  ThrValSerArgSerMetGlyProArgSerGlyValHisGlnArgAlaSerArgIlePro  40
```

RESULT 4
US-60-213-845-350
; Sequence 350, Application US/60213845
; GENERAL INFORMATION:
; APPLICANT: Beasley, Ellen
; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000706
; CURRENT APPLICATION NUMBER: US/60/213,845
; CURRENT FILING DATE: 2000-06-23
; NUMBER OF SEQ ID NOS: 795
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 350
; LENGTH: 236
; TYPE: PRT
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(236)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-60-213-845-350

Alignment Scores:
Pred. No.:
Score: 2.69e-92   Length: 236
                       1078.00   Matches: 236
Percent Similarity: 99.53%   Conservative: 214
Best Local Similarity: 99.53%   Mismatches: 1
Query Match: 39.03%   Indels: 0
DB: 27   Gaps: 0

US-10-014-896-1 (1-1509) x US-60-213-845-350 (1-236)

RESULT 5

US-09-902-540-10715

Sequence 10715, Application US/09902540

GENERAL INFORMATION:

APPLICANT: Goldman, Barry S.
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Wiegand, Roger C.

TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof

FILE REFERENCE: 38-10(15849)B

CURRENT APPLICATION NUMBER: US/09/902,540
CURRENT FILING DATE: 2001-07-10
PRIOR APPLICATION NUMBER: 60/217,883
PRIOR FILING DATE: 2000-07-10
NUMBER OF SEQ ID NOS: 16825
SEQ ID NO 10715

LENGTH: 408
TYPE: PRT
ORGANISM: Myxococcus xanthus

US-09-902-540-10715

Alignment Scores:

Pred. No.:	
Score:	796.50
Percent Similarity:	60.93%
Best Local Similarity:	41.77%
Query Match:	28.84%

Length:	408
Matches:	170
Conservative:	78
Mismatches:	147
Indels:	12
Gaps:	7

US-10-014-896-1 (1-1509) x US-09-902-540-10715 (1-408)

Qy 859 GGAAGCGGGGACAGTGGTCGTATTGCAGCAACTGGCAAAATGAG 903
        ||||||||||||||||||||||||||||||||||||||||||
Db 202 GlySerGlyThrValValThrValLeuGlnGlnLeuAlaAlaAsnGlu 216

Qy 264 TACAGTGGTCAGCACCAGCGTTTATCCAGCATGAGAAGAGTATAGCCACCTGTT 323
        |||| ||| ||| ::||||
Db   5 TyrProArgGlnHisGlnAla-LeuLysArgGlyProValGlyAlaHisSerValLeuTy 24

Qy 324 CACTATCCAAGGCCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTGATGT 383
        :||| ||||||| |||||||||| |||||||||||| |||||||:: |||| |||
Db   24 rThrIleThrGlyThrAspAlaSerLeuArgProAlaLeuLeuLeuGlyHisHisLeuAspVa 44

Qy 384 GGTG--------CCTGCCCGTGAAGAGGGCTGGGAGAGGTGCCCCCAATTCTCTGGGGTTGGA 434
        |||| ||| ||| :||||
Db   44 IValProValGluProGlyThrGluAlaSerTrpThrHisProProTyrSerGlyLeuVa 64

Qy 435 GCCGTGATGGCGTCATCATATGGTGTCGGGCGACACACTGGACACAAGAACTCTGTGATGGCATT 494
        ||||||||||||||||||||||
Db   64 lAlaAspGlyTyrValTrpGlyArgGlyAlaLeuAspAlaProLysGlySerValPheGlyIl 84

Qy 495 ACTGCAGGCCTTGGAGCTCCTGGTCGATCAGGAAGTACATCCCCGGAAGATCTTTCTCAT 554
        ::|||||||||||||||||| :: |||||| ::||||| ::
Db   84 eLeuGluSerValGlnAlaLeuLeuAlaLeuAlaGlyPheGlnProLysArgThrValLeuIle 104

Qy 555 TTCTCTGGGCCATGATGAGGAGTCATCAGGG---ACAGGGCTCAGAGGAATCAGCCCT 611
        |||||| ||| ||||||||||||||| |||||||||||||||||||
Db 104 uAlaPheGlyGlyAspGluGlyValGlyGlyArgGlyAlaGluAlaMetAlaLysIle 124

Qy 612 GCTACAGTCAAAGGGGGCGGGTCCAGCTAGCCTTCATTGTGACAGGAGGGGCTTCAACTTGGA 671
        ||||||||| ||||||||||| :||||||||||:: |||| |::
Db 124 uLeuArgGlyArgGlyValThrLeuGluSerValLeuAspGlyGlyGlyMetIleValSe 144

Qy 672 TGATTCATTCCTAACTTCAAGAAGCCCATCGCCTTGATTGCAGTCTCAGAGAAGGGTTC 731
        :||| |||||||||||||| |||| ||||||||||||||||||||||||||||
Db 144 rGlyThrValProGlyValAlaSerProValAlaLeuValGlyValSerGlyLysGlyPh 164

Qy 732 CATGAACCTCATGCTGCAGTAAACATGACTTCAGGCACTCTTCAGCTCCTCCAAAGGA 791
        ||||||| ||| ||||||||||||||||||||||||||||:: |||
Db 164 eAlaSerAlaLeuLeuValAlaAspGlyLysGlyHisSerSerMetProProProGl 184

Qy 792 GACAAGCAATTGGCATCCTTGCAGCTGCTGTCAGCCGATTGGAGCAGAGACCAATGCCTAT 851
        :||::||||||||||:: :||||||||||||||||||||||||||
Db 184 nThrAlaValAlaGlyValLeuSerArgAlaIleSerArgAlaLeuGluAspAlaProMetProAl 204

Db 184 nThrAlaValAlaGlyValLeuSerArgAlaIleSerArgAlaLeuGluAspAlaProMetProAl 204
        || |||:: ||||
Qy 852 CATATTTGGAAGCGGGGACAGTGGGTGACTGTATTGCAGCAACTGGCAAAT---GAGTTTCC 908
        ||| |||

Db 204 aLysLeuArgGlyGlySer------ArgAlaLeuPheGluProAlaGlyLeuProGluMetGl 222
        ||||
Qy 909 CTTCCCTGTCAATATAATCTGAGCAACCATTGGTAATTTGAACCACTTATAAGCAGGTT 968
        ||| :: :: ||:||| ||||||||||||||||||||

Db 222 yPheGlyMetArgThrLeuPheLeuAlaAlaSnLeuTrpLeuProPheGluProLeuLeuArgGl 242
        ||||||||||||||
Qy 969 TATGGAGAGAAATCCCTTAACCAATCCAATAATCAGGAGCAACACCAGGACACTCACCATATT 1028

Db 242 nLeuThrAlaLysAlaThrThrAsnAlaLeuAlaValArgThrThrThrAlaAlaThrMetPh 262
        :: ||||| ||||||||:||| ||||||||||||||||||
Qy 1029 CAAAGCAGGGGTCAAGTTCAAAGTGTCATCCCCCAGTGGCCCCCAGTCACAGTTCCG 1088

Db 262 eGluGlySerGlnArgAspAsnValLeuProAlaArgGlyAlaValValAlaAsnPheAr 282
        ||||||||||||||||:: ||||||||||||||||||
Qy 1089 GATTCACCCTGGACAGAGCAGTCCAAGAGGTCCTAGAACTCACGAAGAACATTGTGGGCTGA 1148

Db 282 gIleLeuProGlyAspSerValAlaGlyValValLeuGluGluHisValValValAspAs 302
        |||| ||||:||||
Qy 1149 TAACAGAGTCCCAGTTCCATGTGTTGAGTGCCCTTT---GACCCCCTCCCCGTTCAGCCCTTC 1205

Db 302 pProArgValLysValArgThrLeuGlyPheIleSerGlyGluProSerProValSerArgMe 322
        ||| :: ||||||||
Qy 1206 TGATGACAACAAGGGCCTTGGGCTACCCAGCTGTCCTCCGGACCAGCCTTACAGTCCGTTCCCGGA 1265

Db 322 tAspSerGluAla-------TrpSerGlnLeuGlnAsnArgSerValArgGlnValPheProAs 340
        ||| :: ||||
Qy 1266 AGTCAATAATTACTGCCCCAGTTACTTCTTATTGGGCACACACAGCAGCGGATTCTTTACAA 1325

Db 340 pVal---ValAlaAlaProTyrLeuMetLeuGlyAlaThrAspSerArgTyrPheThrGlI 359
        :||| |||||||||:: ||||||
Qy 1326 CCTCACCACTGGCATCTACAGTTCTACCCCATCATACATACAGCCTTGAAGACTTCAAAACG 1385

Db 359 yLeuSerGluAsnValThrArgPheMetProLeuArgLeuAspGlyAlaAspLeuSerArA 379
        ||| :: ||||
Qy 1386 CATCCCATGGAGTCAACGAGAGAAAATCTCAGTCCAAGCCTATAGAGACCCAAGTGAAATTCAT 1445

Db 379 gLeuHisGlyLysAspGluArgValSerValLysGlyThrValAlaAspAlaValArgPheTy 399
        ||| :: ||||||||||
Qy 1446 CTTTGAGTTGATTCAGAAT 1464

Db 399 rAlaGlnThrValArgAsn 405

RESULT 6

PCT-US02-09921-1008

Sequence 1008, Application PC/TUS0209921

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: DAFFO, Abel
APPLICANT: JONES, Anissa L.
APPLICANT: TRAN, Alanna-Phung B.
APPLICANT: DAHL, Christopher R.
APPLICANT: CHINN, Darryl
APPLICANT: GIETZEN, Joyce
APPLICANT: DUFOUR, Gerard E.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: YU, Jimmy Y.
APPLICANT: TUASON, Olivia
APPLICANT: YAP, Pierre E.
APPLICANT: AMSHEY, Stefan R.
APPLICANT: DAUGHERTY, Sean C.
APPLICANT: DAM, Tam C.
APPLICANT: LIU, Tommy F.
APPLICANT: NGUYEN, Duy-Viet An
APPLICANT: KLEEFELD, Yael
APPLICANT: GERSTIN JR., Edward H.
APPLICANT: PERALTA, Careyna H.
APPLICANT: DAVID, Marie H.
APPLICANT: LEWIS, Samantha A.
APPLICANT: CHEN, Alice J.

; APPLICANT: PANZER, Scott R.
; APPLICANT: HARRIS, Bernard
; APPLICANT: FLORES, Vincent
; APPLICANT: MARWAHA, Rakesh
; APPLICANT: LO, Audrey
; APPLICANT: LAN, Ruth Y.
; APPLICANT: URASHKA, Michael
; TITLE OF INVENTION: SECRETORY MOLECULES
; FILE REFERENCE: PF-1232 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/09921
; CURRENT FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
; PRIOR FILING DATE: 2001-03-29; 60/299,428; 60/300,001; 60/299,776
; PRIOR FILING DATE: 2001-05-17; 2001-03-29; 2001-06-19; 2001-05-16; 2001-05-17;
; NUMBER OF SEQ ID NOS: 1146
; PRIOR APPLICATION NUMBER: 2001-06-19; 2001-06-20
; SEQ ID NO 1008
; SOFTWARE: PERL Program
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: LG:369881.5.orf1.2001JUN22
PCT-US02-09921-1008

Alignment Scores:
Pred. No.:
Score: 794.00
Percent Similarity: 100.00%
Best Local Similarity: 100.00%
Query Match: 28.75%

Length:	154	
Matches:	154	
Conservative:	0	
Mismatches:	0	
Indels:	0	
Gaps:	0	

US-10-014-896-1 (1-1509) x PCT-US02-09921-1008 (1-154)

```
Qy  1045 TTCAAATGTCATCCCCCCAGTGGCCCAGGCCACAGTCAAACTTCCGGATTCACCCTGGACAG 1104
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 PheAsnValIleProProValAlaGlnAlaThrValAsnPheArgIleHisProGlyGln 20

Qy  1105 ACAGTCCAAGAGAGGTCCTAGAACTCACGAAGAAACATTGTGGCTGATAACAGAGTCCAGTTC 1164
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    21 ThrValGlnGluValLeuGluLeuThrLysAsnIleValAlaAspAsnArgValGlnPhe 40

Qy  1165 CATGTGTTGAGTGGCCTTTGACCCCCTCCCGTCACCCCTTCTTGATGACAAGGCCTTGGGC 1224
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    41 HisValLeuSerAlaPheAspProLeuProValSerProSerProSerAspLysAlaLeuGly 60

Qy  1225 TACCAGTGCTCCGCCAGACCGTACAGTCCGTCTTCCCGGAAGTCAATAATTACTGCCCCA 1284
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 TyrGlnLeuLeuArgGlnThrValGlnSerValPheProGluValAsnIleThrAlaPro 80

Qy  1285 GTTACTTCTTATTGGCAACACAGAGCACGCGATTCTTTACAAACCTCCACCACTGGCATCTAC 1344
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    81 ValThrSerIleGlyAsnThrAspSerArgPhePheThrAsnLeuHisThrGlyIleTyr 100

Qy  1345 AGGTTCTCTACCCCCATCTACAATACAGCCTGAAGACTTCAAACGCATCCATGGAGTCAACGAG 1404
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ArgPheTyrProIleTyrIleGlnProGluAspPheLysArgIleHisGlyValAsnGlu 120

Qy  1405 AAAATCTCAGTCCAAGCCTATAGAGACCCAAGTGAAAATTCATCTTTGAGTTGATTCAGAAT 1464
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LysIleSerValGlnAlaTyrGluThrGlnValLysPheIlePheGluLeuIleGlnAsn 140

Qy  1465 GCTGACACGAGCCAGGAGGAGCCAGTTTCCTCACCTGCACAAACTG 1506
         ||||||||||||||||||||||||||||||||||||||||||||||
Db   141 AlaAspThrAspGlnGluProValSerHisLeuHisLysLeu 154
```

; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
; FILE REFERENCE: 806CIP/PCT
; CURRENT APPLICATION NUMBER: PCT/US02/25485
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: PCT/US02/25485
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,261
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2986
; SOFTWARE: Pt_FL_genes Version 5.0
; SEQ ID NO 2574
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
PCT-US02-25485-2574

Alignment Scores:
Pred. No.:
Score: 562.50
Percent Similarity: 79.72%
Best Local Similarity: 79.72%
Query Match: 20.37%

Length:	120	
Matches:	114	
Conservative:	0	
Mismatches:	0	
Indels:	29	
Gaps:	1	

US-10-014-896-1 (1-1509) x PCT-US02-25485-2574 (1-120)

```
Qy     1 ATGGCTCAGCGCGTGCGCTTTGCGTGCTGTGGCCCTGTGGGCTATGCCGCTTCCTAGTTTTCCT 60
         |||||||||||||||||||||||||||||||||||||||||||||
Db     7 MetAlaGlnArgCysValCysValLeuAlaLeuValAlaAlaMetLeuLeuLeuValPhePro 26

Qy    61 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGGAGCATCAAAGGGCGTCGCGAATCCCT 120
         ||||||||||||||||||||||||||||||||||||||||||||||
Db    27 ThrValSerArgSerMetGlyProArgSerGlyLeuHisGlnArgAlaSerArgIlePro 46

Qy   121 TCTCAGTTCACCAAAAGGAACGCGCTTCGGCGATGAAAAGACGGCGCGTGAAAGGTGCCATCCAG 180
         |||||||||||||||||||||||||||||||||||||||||||
Db    47 SerGlnPheSerLysGluGlyArgValAlaMetLysGluAlaLeu- 61

Qy   181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATAACTACAGCGCCTGGCTGAGTCGGA 240
         |||||||||||||||||||||||||||||||||||||||||||||
Db    61 ------------------------------------------------ 61

Qy   241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGGACCACCAGCTTTATCCAGCATGAAGTC 300
         |||||||||||||||||||||||||||||||||
Db    61 LysValPheProThrValValSerThrSerPheIleGlnHisGluVal 77

Qy   301 GTGGAAGAGTATAGCCACCGTTCACTATCCAAGCCTCGGACACCCCAGCGTTGCAGCCCTAC 360
         |||||||||||||||||||||||||||||||||||||||||||||
Db    78 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr 97

Qy   361 CTGCTGATGGGCTCACTTTGATGTGGTGGCTGCCTGCCCCTGAACAAAGGCGTGGGAGGTGCCCCA 420
         |||||||||||||||||||||||||||||||||||||||||||
```

; GENERAL INFORMATION:
; APPLICANT: Tang, X. Tom
RESULT 7
PCT-US02-25485-2574
; Sequence 2574, Application PC/tUS025485
; GENERAL INFORMATION:
; APPLICANT: Tang, X. Tom

Db   98 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro 117

Qy 421 TTCTCTGGG 429
        |||||||||
Db 118 PheSerGly 120

RESULT 8
US-60-311-261-2481
; Sequence 2481, Application US/60311261
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Weng, Gezhi
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
; FILE REFERENCE: 806
; CURRENT APPLICATION NUMBER: US/60/311,261
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/649,167
; PRIOR FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 09/552,929
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-04-18
; PRIOR APPLICATION NUMBER: US 09/770,160
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 09/540,217
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/577,408
; PRIOR FILING DATE: 2000-05-18
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: Pt_FL_genes Version 5.0
; SEQ ID NO 2481
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-311-261-2481

Alignment Scores:
Pred. No.:
Score: 1.02e-43
Percent Similarity: 562.50
Best Local Similarity: 79.72%
Query Match: 20.37%
DB:

		Length:	120
		Matches:	114
		Conservative:	0
		Mismatches:	0
		Indels:	29
		Gaps:	1

US-10-014-896-1 (1-1509) x US-60-311-261-2481 (1-120)

Qy   1 ATGGCTCAGCGGTGCGTTCGGTTGCTGCTCGTGCTGCTCCTGGTCGCTCTATGCTGGCTCTATGCTGCTCCTAGTTTTTCCT 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   7 MetAlaGlnArgCysValCysValLeuLeuAlaLeuAlaMetLeuLeuLeuValPhePro 26

Qy 61 ACCGTCTCCAGATCGATGGGCCCGAGAGAGCGGGGAGCAATCAAAAGGGCCTCGGCGAAATCCCT 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 27 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 46

Qy 121 TCTCAGTTCAGCAAGAGAAACGGCGTCGGGATGAAAAGAGGAGGCGTTGAAAAGGTGCCATCCAG 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   47 SerGlnPheSerLysGluGluArgValAlaMetLysGluAlaLeu 61

Qy 181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGGA 240
       ||||||||||||||||||||||||||||
Db   62 -------------LysValPheProThrValValSerThrSerPheIleGluHisGluVal 77

Qy 241 AAATACATTCATTAAAGTCTTTCCTACAGTGGTCACCACCAGCTTTATTCAGCATGAAGTC 300
       |||||||||||||||||||||||||||||||||||||||||
Db   61 --- 61

Qy 301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTGGAGACCCCAGCTTGCAGCCCTAC 360
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr 97

Qy 361 CTGCTGGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCCA 420
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   98 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro 117

Qy 421 TTCTCTGGG 429
       |||||||||
Db 118 PheSerGly 120

RESULT 9
US-10-179-131-8920
; Sequence 8920, Application US/10179131
; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 8920
; LENGTH: 581
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-8920

Alignment Scores:
Pred. No.:
Score: 9.31e-41
                535.00
Percent Similarity: 46.65%
Best Local Similarity: 31.36%
Query Match: 19.37%
DB:

		Length:	581
		Matches:	164
		Conservative:	80
		Mismatches:	197
		Indels:	82
		Gaps:	20

US-10-014-896-1 (1-1509) x US-10-179-131-8920 (1-581)

Qy 118 CCTTCTCAGTTCAGCAAAGAGGAACGGCGTCGGGATG----------------- 153
       ||| ::: ||||:::|
Db   64 ProGluSerTyrTyrLysAspAspAsnSerThrValLeuGluIleLeuHisAspLysLysTyr 83

Qy 154 ---AAAAGAGGCGCTGAAA---------GGTGCCATCCAGATTCCAACAGTGACTTTTAGC 201
       ||:|||||||||:::
Db   84 LysLysGluSerIleLysArgLeuAlaGlyAlaIleGlnValAspThrGlnValPheAsp 103

Qy 202 TCTGAGAAGTCCAATACTACAGCC---------------CTGGCTGGAGTTCGGAAAA 243
       ::: ||| ||:::|||
Db 104 LysGlnProAlaValAspAspAlaProGlnValTrpLysLysPheHisAsp 123

Qy 244 TACAATCAATAAAAGTCTTTCCTACAGTGGTCACCACCAGCTTTATCCACCATGAAGTCGTG 303
       ||::: |||||| |||
Db 124 TyrLeuGluGlnThrPheProLeuValTyrLysAsn-----LeuLysValThrLysVal 141

```
Qy    304 GAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTG 363
                |||      ::::::      :::|||||||||   ||||||:::|||   |||
Db    142 AsnThrTyrGlyLeuValTyrHisTrpLysGlySerAspLysSerLeuLysProValLeu 161

Qy    364 CTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAA---------GGCTGGGAGGTG 414
          |||   ||||||   |||   ||||||      ::::::            |||
Db    162 LeuThrAlaHisGlnAspThrValProValGlnLysAspThrLeuLysAspTrpThrTyr 181

Qy    415 CCCCCATTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGGCACACTGGACGAC 474
          |||||||||   |||                  |||||||||||||||      |||
Db    182 ProProPheGluGlyHisTyrAspGlyGluTyrIleTyrGlyArgGlyAlaAlaAspCys 201

Qy    475 AAGAACTCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATC 534
          ||||||   ::::::|||:::|||:::   |||||||||||||||   :::   |||
Db    202 LysAsnValLeuIleAlaIleLeuGluThrLeuGluLeuLeuLeuAlaLysGlyTyrGln 221

Qy    535 CCCCGAAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTCATCAGGG---ACAGGG 591
          |||:::||||||         :::   |||   |||||||||:::||||||      |||
Db    222 ProLysArgSerIleLeuAlaAlaPheGlyPheAspGluGluAlaSerGlyTyrHisGly 241

Qy    592 GCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGCGTCCAG------CTAGCCTTCATT 645
          |||      |||         |||:::            |||      :::      :::
Db    242 AlaAlaHisIleGlyLysTyrLeuGluGluThrPheGlyGlnAspSerValTyrAlaLeu 261

Qy    646 GTGGACGAGGGGGGCTTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCC 705
          :::|||||||||      :::            :::   |||            :::|||
Db    262 IleAspGluGlyAlaGlyLeuThrValGlnGluLeuThrAsn------ThrIleValAla 279

Qy    706 TTGATTGCAGTCTCAGAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCA 765
          |||         :::|||||||||   :::::::::   :::::::::
Db    280 LeuProGlyThrAlaGluLysGlyTyrValAspIleGlnValGluLeuThrThrProGly 299

Qy    766 GGCCACTCTTCAGCTCCTCCAAAGGAGACAAGCATTGGCATCCTTGCAGCTGCTGTCAGC 825
          ||||||||||||   ||||||      |||||||||||||||::::::
Db    300 GlyHisSerSerIleProProAspProThrSerIleGlyIleIleSerGluLeuGlyTyr 319

Qy    826 CGATTGGAGCAGACACCA---ATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTA 882
             :::|||:::   |||      |||::::::               ::::::
Db    320 IleIleGluLysAspProTyrSerProLeuLeuProProGluAsnProIleLeuAsnPhe 339

Qy    883 TTGCAGCAACTGGCA---------------AATGAGTTTCCC------TTCCCTGTCAAT 921
             |||   ||||||               |||      |||      |||
Db    340 AlaGlnCysLeuAlaLeuHisAspProLysAsnAsnIleProSerSerPheLysLysAla 359

Qy    922 ATAATC------------------CTGAGCAACCCATGGCTATTTGAACCACTTATAAGC 963
          |||:::                     :::|||      |||   |||   :::
Db    360 IleLeuArgAlaGlyTyrAspLysPheAlaAsnSerLysLeuValGluGlyIle------ 377

Qy    964 AGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTCACC 1023
                      :::|||   ||||||      :::|||||||||:::   ||||||
Db    378 ---------SerLysAsnArgLeuThrLysTyrLeuIleArgThrSerGlnAlaLeuAsp 394

Qy   1024 ATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGCCACAGTCAAC 1083
          |||         |||   |||   |||   :::|||         :::      ||||||
Db    395 IleIleAsnGlyGlyGluLysAlaAsnAlaLeuProGluHisValLysLeuLeuValAsn 414

Qy   1084 TTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAA---CTCACGAAGAACATT 1140
             |||:::      |||   :::|||   ||||||   |||               :::
Db    415 HisArgValAlaIleGlyThrSerValAlaGluValGlnGluHisPheValSerArgVal 434

Qy   1141 GTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTT------------------ 1182
          |||         :::               :::|||||||||
Db    435 ValGluValAlaLysArgHisGlyLeuSerValSerAlaPheGlyLysAspValLeuLys 454

Qy   1183 ---------------------------------------------------GACCCCCTC 1191
                                                             :::
Db    455 ValLysAsnAspSerGlyLeuPheAsnValThrAsnPheAlaGlyPheLeuAsnAlaAla 474

Qy   1192 CCCGTCAGCCCTTCTGATGACAAGGCCTTGGGCTAC------CAGCTGCTCCGCCAGACC 1245
          ||||||:::|||::::::|||            |||         :::   |||
Db    475 ProValThrProThrAsnAspThrValTrpGluTyrLeuSerGlyValThrArgHisVal 494

Qy   1246 GTACAGTCC---GTCTTCCCGGAAGTCAAT---------ATTACTGCCCCAGTTACTTCT 1293
             :::      ||||||||||||:::            ||||||||||||
Db    495 TyrGluAspLeuValPheProGluIleLysTyrProValIleThrAlaProAlaIleMet 514

Qy   1294 ATTGGCAACACAGACAGCCGATTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTAC 1353
             ||||||||||||:::|||   :::   |||||||||      |||:::||||||
Db    515 ThrGlyAsnThrAspThrArgHisTyrTrpAsnLeuThrArgAsnIlePheArgPheThr 534

Qy   1354 CCCATCTACATACAGCCTGAAGACTTC---------AAACGCATCCATGGAGTCAACGAG 1404
          |||   :::|||         ||||||               ||||||   |||:::|||
Db    535 ProAlaPheIle------GlyAspPheIleGlyGluThrHisIleHisSerValAspGlu 552

Qy   1405 AAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAAATTCATCTTTGAGTTGATTCAGAAT 1464
          |||:::         ::::::               :::   :::|||   ||||||
Db    553 LysLeuProPheAspSerHisLeuGlnLeuGlnAlaTrpPheTyrGluTyrIleGlnAla 572

Qy   1465 GCTGACACA 1473
             |||:::
Db    573 IleAspSer 575

RESULT 10
US-09-248-796-17521
; Sequence 17521, Application US/09248796
; GENERAL INFORMATION:
;  APPLICANT: Keith Weinstock et al
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
;  TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: 107196.132
;  CURRENT APPLICATION NUMBER: US/09/248,796
;  CURRENT FILING DATE: 1999-02-12
;  NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17521
;   LENGTH: 616
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-09-248-796-17521

Alignment Scores:
Pred. No.:              1.85e-40        Length:       616
Score:                  532.00          Matches:      163
Percent Similarity:     46.65%          Conservative: 81
Best Local Similarity:  31.17%          Mismatches:   197
Query Match:            19.26%          Indels:       82
DB:                     16              Gaps:         20

US-10-014-896-1 (1-1509) x US-09-248-796-17521 (1-616)

Qy    118 CCTTCTCAGTTCAGCAAAGAGGAACGCGTCGCGATG------------------------ 153
          |||      :::   |||:::            :::
Db     99 ProGluSerTyrTyrLysAspAsnSerThrValLeuGluIleLeuHisAspLysLysTyr 118

Qy    154 ---AAAGAGGCGCTGAAA---------GGTGCCATCCAGATTCCAACAGTGACTTTTAGC 201
             ||||||::::::|||         ||||||||||||:::   |||      |||
Db    119 LysLysGluSerIleLysArgLeuAlaGlyAlaIleGlnValAspThrGlnValPheAsp 138

Qy    202 TCTGAGAAGTCCAATACTACAGCC------------------CTGGCTGAGTTCGGAAAA 243
             :::   :::         |||                     |||:::|||
Db    139 LysGlnProAlaValAspAspAlaProGlnValTrpAlaLysPheAlaLysPheHisAsp 158

Qy    244 TACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTCGTG 303
          |||:::   :::   ||||||   |||               ::::::         |||
Db    159 TyrLeuGluGlnThrPheProLeuValTyrLysAsn------LeuLysValThrLysVal 176

Qy    304 GAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTG 363
                |||      ::::::      :::|||||||||   ||||||:::|||   |||
Db    177 AsnThrTyrGlyLeuValTyrHisTrpLysGlySerAspLysSerLeuLysProValLeu 196

Qy    364 CTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAA---------GGCTGGGAGGTG 414
          |||   ||||||   |||   ||||||      ::::::            |||
Db    197 LeuThrAlaHisGlnAspThrValProValGlnLysAspThrLeuLysAspTrpThrTyr 216

Qy    415 CCCCCATTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGGCACACTGGACGAC 474
          |||||||||   |||                  |||||||||||||||      |||
Db    217 ProProPheGluGlyHisTyrAspGlyGluTyrIleTyrGlyArgGlyAlaAlaAspCys 236
```

RESULT 11
US-60-096-409-17521
Sequence 17521, Application US/60096409A
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBI
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-016P
; CURRENT APPLICATION NUMBER: US/60/096,409A
; CURRENT FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28206
; SEQ ID NO 17521
;   LENGTH: 616
;   TYPE: PRT
;   ORGANISM: Candida albicans
US-60-096-409-17521

Alignment Scores:
Pred. No.:
Score: 1.85e-40
Percent Similarity: 532.00   Length: 616
Best Local Similarity: 46.6%   Matches: 163
Query Match: 31.17%   Conservative: 81
DB: 19.26%   Mismatches: 197
                              27   Indels: 82
                                              Gaps: 20

US-10-014-896-1 (1-1509) x US-60-096-409-17521 (1-616)

Qy   118 CCTTCTCAGTTCAGCAAAGAGAGGAAACCGTCGCGATG--------------------- 153
                ::: ||||||||||||||||||| ::::

Db   99   ProGluSerTyrTyrLysAspAsnSerThrValLeuGluIleLeuHisAspLysTyr 118

Qy   154 ---AAAGAGGGCGCTGAAA----------GGTGCCATCCAGAATCCAACAGTGACTTTTAGC 201
              ||||| ||||| | ||||||||||||||| ||||||||| |||

Db   119 LysLysLysSerIleLysArgLeuAlaGlyAlaIleGlnValAlaAspThrGlnValPheAsp 138

Qy   202 TCTGAGAAGTCCAAATACTACAGCC------------------CTGGCTGAGTTCGGAAAA 243
              ||| : ||||:|||||||||||

Db   139 LysGlnProAlaValAlaAspAspAlaProGlnValTrpAlaLysPheAlaLysPheHisAsp 158

Qy   244 TACATTCATAAAGTCTTTCCTACAGTGGTCACGCACCACCAGCTTTATCCAGAGCAAGTCGTG 303
              || ||| :::| || ::: |||||||| :::

Db   159 TyrIleGluGlnThrPhePheProLeuValTyrLysAsn------LeuLysValThrLysVal 176

Qy   304 GAAGAGTATAGCACACCTGTTCACTATCCAAAGGCTCGGACCCCAGCTTGCAGCCTACCTG 363
              |||||||||||||||| ||||| |:|||||||||||:|||||||||||||||||||

Db   177 AsnThrTyrGlyLeuValTyrHisThrProLysGlySerAspLysSerLeuLysProValLeu 196

Qy   364 CTGATGGGTCACTTGATGTGGTGCCTCGCCCCTGCCCCCTGAAGAA----------GGCTGGGAGGTG 414
              ||||| :::::|:||| |||| ::::   :::

Db   197 LeuThrLeuAlaHisGlyLnAspThrValProValGlnLysAspThrTrpThrTyr 216

Qy   415 CCCCCATTCTCTCGGGTTGGAGCGTTCGATGGTGCATCTARCGTGCGGGGCACACTGGACGAC 474
              ||||||||||| ||||||||||||||||||||||| |||

Db   217 ProProPheGluGlyHisThrTyrAspGlyIleTyrIleGlyArgGlyIleAlaLeuAspAspCys 236

Qy   475 AAGAACTCTGTGATGGGCAATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATC 534
                ||||| :::||||||||||||||||| ::::||||||||||||||||||

Db   237 LysAsnValLeuIleLeuAlaIleLeuGluThrLeuLeuLeuAlaLysGlyTyrGln 256

Qy   535 CCCCGAAGATCTTTCTTCATTCTCTGGGCCCATGATGAGGAGTCATCAGG---ACAGGG 591
              |||||||||||||||||||||| ||| ||||| ||||||| ::

Db   257 ProLysArgSerIleLeuLeuAlaAlaAlaPheGlyPheAspGluThrSerGlyTyrHisGly 276

Qy   592 GCTCAGAGGATCTCAGACCTGTCACAGTCAAAGGGCGTCCAG-----CTAGCCTTCATT 645
              ||| :::|| :::   ::

Db   277 AlaAlaHisIleGlyLysTyrLeuGluThrPheGlyLnAspSerValTyrArgLeu 296

Qy   646 GTGGACGAGGGGGGCTTCATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGCC 705
              :::||||||||||||||||||||||||||||||||||||||||||||||||

Db   297 IleAspGluGlyAlaGlyLeuThrValGlGluLeuThrAsn------ThrIleValAla 314

Qy   706 TTGATTGCAGTCTCAGAGAAGGGTTCCATGAAACTCATGCTGCAAGTAAACATGACTTCA 765
              ||| ||||||||||||||||||||||||||||| :::

Db   315 LeuProGlyThrAlaGluLysGlyTyrValAspIleGlnValGlyLeuThrThrProGly 334

Qy   766 GGCCCACTCTTCAGCTCCTCCAAAAGCAGACAAAGCAATTGGGCAATCCTTGCCGTGCTGTCAGC 825
              ||||||||||||||||||||||||| ||||||||||||||||| :::

Db   335 GlyHisSerSerIlePGoProGluAspProThrSerIleGlyIlaIleIleSerGluGluLeuGlyTyr 354

Qy   883 TTGCAGAACTGGCA----------AATGAGTTTCCC---------TTCCCTGTCAAT 921
              ||| |||||||| |||||||| |||||||||||

Db   375 AlaGlnCysLeuAlaLeuHisSerAspProSerSerPheLysLysAla 394

Qy   964 AGGTTTATTGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTCACC 1023
              :::|||||||||||||||||||||| ||| |||||||

Db   413 ---SerLysAsnArgLeuThrLysTyrLeuLeuArgThrSerGlnAlaLeuAsp 429

Qy   1024 ATATTCAAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCCAGGGCCCCACAGTCAAC 1083
              :::||||||||||||||||||||||||

Db   430 IleIleAsnGlyGlyGlyLysSerAlaAsnAlaLeuProGlnHisValLysLeuValAsn 449

Qy   1084 TTCCGGATTCACCCTGGACAGCAGTCCAAGAGGTCCTAGAA---CTCACGAAGAACATT 1140
              ||| :::|||||||||||||||||||||||||||| ::

Db   450 HisArgValAlaIleGlyThrSerValAlaGluValGlnGluHisPheValSerArgVal 469

Qy   1141 GTGGCTGTGATAACAGAGTCCAGTTCCATCTGTGTTGAGTGCCTTT-------------- 1182
              :::||||||||||||||||||||||||||||||||||||||

Db   470 ValGluValAlaLeuAlaArgHisGlyLeuSerValAlaPheGlyLeuLysAspValLysLys 489

Qy   1183 ----------------------------------GACCCCCTC 1191
                                                |||

Db   490 ValLysAsnAspSerGlyLeuPheGluAsnAlaGlyPheLeuAsnAlaAla 509

Qy   1192 CCCGGTCAGCCCTTCGATGACAAGGCCTTGGGGTAC----CAGCTGCTCCGGCCAGACC 1245
              |||||||||||||||||||||||||||||||| :::

Db   510 ProValThrProThrArgAsnAspThrValTrpGluThrTyrLeuSerGlyValThrLysHisVal 529

Qy   1246 GTACAGTCC--GTCTTCCGGAAGTCAAT---------ATTACTGCCCCAGTTACTTCT 1293
              |||||||| |||||||||||||| ||||

Db   530 TyrLeuAspLeuValPheProGlyIleLeuLysTyrProValIleGluThrAlaProIleAlaIleMet 549

Qy   1294 ATGGGGCCACACACAGCCGGATTCTTTACAACAACCTGGCACTACAGGTTGTAC 1353
              |||||||||||| :::|||||||||||||||||| :::

Db   550 ThrGlyAlaThrHisAspThrArgHisSerTyrArgAsnIlePheArgPheThr 569

Qy   1354 CCCATCTACATACAGCCTGAAGAGACTTC-------AAACGCAATCCATGGAGTCAACGAG 1404
              ||| ||| :::|||||||| |||||||||||| ::::

Db   570 ProAlaPheIle------GlyAspPheIleGlyGluThrHisHisSerValAspGlu 587

Qy   1405 AAAATCTCAGTCCAAGCCCTATGAGACCCCAAGTGAAATTCATCTTTGAGTGTGGTTGGAATT 1464
              :::::   ::: :::|||||| ::::||| :||||

Db   588 LysLeuProPheAspSerHisLeuGlnLeuGlnIleAlaThrTrpPheGlyLnAlaIleTrpPheGluTyrLys 607

Qy   1465 GCTGGACACA 1473
              ||| ||:::

Db   608 IleAspSer 610

; GENERAL INFORMATION:
; APPLICANT: HARE, ROBERTA S.
; APPLICANT: SHAW, KAREN J.
; APPLICANT: SHIMER JR., GEORGE H.
; APPLICANT: KESSLER, MARCO
; APPLICANT: NOLLING, JORK
; APPLICANT: ZENG, QIANDONG
; APPLICANT: GREENE, JONATHAN R.
; TITLE OF INVENTION: CANDIDA ALBICANS NUCLEIC ACIDS AND POLYPEPTIDES,
; TITLE OF INVENTION: AND USES THEREFOR
; FILE REFERENCE: 2976-4031
; CURRENT APPLICATION NUMBER: US/10/179,131
; CURRENT FILING DATE: 2002-06-21
; NUMBER OF SEQ ID NOS: 10194
; SEQ ID NO 7140
; LENGTH: 623
; TYPE: PRT
; ORGANISM: Candida albicans
US-10-179-131-7140

Alignment Scores:
Pred. No.:
Score: 7.39e-38
Percent Similarity: 504.50
Best Local Similarity: 47.72%
Query Match: 28.51%
DB: 18.27%
                              25

US-10-014-896-1 (1-1509) x US-10-179-131-7140 (1-623)

Length: 623
Matches: 144
Conservative: 97
Mismatches: 195
Indels: 69
Gaps: 17

RESULT 12
US-10-179-131-7140
; Sequence 7140, Application US/10179131

```
Qy    724 AAGGGTTCCATGAACCTCATGCTGCGAAGTAAACATGACTTCAGGCCACTCTTCAGCTCCT 783
              ||||||        ::||| |||      :::||| |||
Db    333 LysGlyHisLeuAsnSerValIleAspLeuPheThrProGlyGlyHisSerValPro 352

Qy    784 CCAAAGGAGGACAAGCATTGGCATCCTTGCAGCTGCTGTCACCGATTGGGAGCAGGACACCA 843
              ||||||        :::|||||||||| |||||:::||||
Db    353 ProArgHisThrSerIleGlyIleMetSerGlnLeuIleThrLysIleGluAspLysGlu 372

Qy    844 ATG---CCTATCATATTTGGAAGGGGACAGTGGTGACTGTGATTGCAGCAACTGCCAAAT 900
              |||||        ::|||||||        ||:|||
Db    373 PheSerProIleLeuThrAsnAsnProValLeuGlyLnLeuTyrCysLeuAla--- 391

Qy    901 GAGTTTCCCTTCCTTCCTTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTT--- 957
              |||        ::||||||        :::|||
Db    392 GluHisSerLysIleLeuAsnLysSerIleLysArgAspSerIleLeuLysAlaGlnIleAsp 411

Qy    958 ------------ATAAGCAGGTTTATGGAGAGAGAAATCCCTTACCAATGCAATA 999
                                   :::||| :::|||
Db    412 GlnAsnAlaAsnGlnAsnArgValValGluTyrLeuSerLysAspAlaGluThrLysTyrLeu 431

Qy   1000 ATCAGGACCACCACGGCCACTCACCATATTCAAAAGCAGGGGTTCAAGTTCAATGTCATCCCC 1059
              ||||||        ||||||||| |||||||        :::||||
Db    432 IleThrThrSerGlnAlaValAspIleIleGlnLysGlnValLysSerAsnAlaLeuPro 451

Qy   1060 CCAGTGGCCCAGGCCAACTTCCGGGATTCACCCTGGACAACAGTCCAAGAGGTC 1119
              ||||||        :::|||
Db    452 GluHisValSerValValAlaSnSerArgIleAlaValGluGlnAsnValAsnThrVal 471

Qy   1120 CTAGAACTCACGAAG------------AACATTGTG 1143
              ||||||
Db    472 ValSerLysLysIleLeuLysAspIleLeuThrIleAlaAspLysPheAsnLeuGlyLeuIle 491

Qy   1144 GCTGATAACAGA------------GTCCAGTTCCATGTGGTTG 1173
              |||||||        :::||| :::|||
Db    492 IleAspAsnArgGlnGluIleIleGluProThrGluHisGlyTyrPheAsnTyrSerLeuIle 511

Qy   1174 AGTGCCCTTTGACCCCCGTCAGCCCTTCTGATGACACAAGGCC------------TTGGGC 1224
              |||||||        :::||| :::|||
Db    512 GluSerLeuGluProAlaProValSerProIleAsnGlySerTrpAsnIlePheGly 531

Qy   1225 TACCAGCTGCTCCCGCCAGACCGTACACGTTCCGGTCTTCCCGGAAGTCAAT------------ATT 1275
              |||        :::|||
Db    532 GlySerLeuArgTyrLeuArgTyrLeuAspLeuIlePheProAspSerAsnAspThrPheIle 551

Qy   1276 ACTGCCCCAGTTACTTCTTATTGGCAACACAGACAGCCGGATTCTTTACAAAACCTCACCACT 1335
              ||||||        ::: :::||||
Db    552 ValAlaProPheLeuSerThrGlyAsnThrAspThrLysSerTyrTrpAspSerLeuThrArg 571

Qy   1336 GGCATCTACAGGTTCTACCCCACATCTACACAGACTTCAAACGACATCCATGGA 1395
              ||||||        :::|||
Db    572 AsnIleTyrArgTyrGlnProSerIleAlaThrLysAsn------SerAsnIleHisSer 589

Qy   1396 GTCAACGAGAAAATCTCAGTCAAGCCTATGAGACCCAAGTGAAAATTCATCTTGAGTTG 1455
              |||        :::||||
Db    590 IleAspGlyLysLeuAspPheGluGlyHisPheHisIleIleAlaPheTyrTyrTyrTyr 609

Qy   1456 ATTCAGAATGCTGAC 1470
              ::::||        |||
Db    610 LeuGlnLeuValAsp 614
```

```
Alignment Scores:
Pred. No.:                          3.09e-33        Length:         93
Score:                              451.00          Matches:        93
Percent Similarity:                 98.94%          Conservative:   0
Best Local Similarity:              98.94%          Mismatches:     0
Query Match:                        16.33%          Indels:         1
DB:                                 13              Gaps:           0

US-10-014-896-1 (1-1509) x US-08-904-468-192 (1-93)

Qy      1 ATGGCTCAGCGGTGCCGTTTGCCGTGCCTGCCCTGGTTGGCCTATGCTGCTCCTAGTTTTCCCT 60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaGlnArgCysValCysValLeuAlaLeuAlaLeuLeuLeuValAlaLeuLeuLeuValPro 20

Qy     61 ACCGTCTCCAGATCGATGGGCCCCGAGGAGCGGGGGAGCATCAAAGGCGTCGGGGAATCCCT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ThrValSerArgSerMetGlyProArgSerGlyIleGluHisGlnArgAlaSerArgIlePro 40

Qy    121 TCTCAGTTCCAGCAAAGAGGAGGAAACGCGTCCGGATGAAAAAGAGCGGCGATGAAAGGTGCCATCCAG 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerGlnPheSerGlyGluGlnArgValAlaMetLysGluLeuLysGlyAlaIleGln 60

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGCCTGAGTTCGGA 240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaLeuAlaGluPheGly 80

Qy    241 AAATACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACCA 280
            |||||||||||||||||||||||||||||||||||||||
Db     81 -AsnThrPheIleLysSerPheLeuGlnTrpSerAlaPro 93
```

; APPLICANT: Duclert, Aymeric
; APPLICANT: Lacroix, Bruno
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins
; FILE REFERENCE: GEN-T119C1
; CURRENT APPLICATION NUMBER: US/09/547,599C
; CURRENT FILING DATE: 2000-04-11
; PRIOR APPLICATION NUMBER: US 08/905,223
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,135
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,051
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,144
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,279
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/904,468
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,134
; PRIOR FILING DATE: 1997-08-01
; PRIOR APPLICATION NUMBER: US 08/905,133
; PRIOR FILING DATE: 1997-08-01
; NUMBER OF SEQ ID NOS: 3475
; SOFTWARE: Patent.pm

; SEQ ID NO 2910
; LENGTH: 93
; TYPE: PRT
; ORGANISM: Homo Sapiens
; FEATURE:
; OTHER INFORMATION: Large Intestine
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -23..-1
; OTHER INFORMATION: Von Heijne matrix score 9.10
; OTHER INFORMATION: seq LVAMLLVFPTVS/RS

US-09-547-599C-2910

Alignment Scores:
Pred. No.:   3.09e-33
Score: 451.00
Percent Similarity: 98.94%   Length: 93
Best Local Similarity: 98.94%   Matches: 93
Query Match: 16.33%   Conservative: 0
DB: Mismatches: 0
                                          Indels: 1
                                          Gaps: 0

US-10-014-896-1 (1-1509) x US-09-547-599C-2910 (1-93)

Qy   1 ATGGCTCAGCGGTGCGTTGCGTTTGCGTGCTGGGCCCTGGTGGTGCTCCTCAGTTTTTCCTT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaGlnArgCysValCysValLeuAlaLeuValAlaLeuLeuLeuLeuValPhePro 20

Qy 61 ACCGTCTCCAGATCGATGGGCCCGAGGAGGGGAGGCATCAAAAGGGCGTCGCCAAATCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 21 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 40

Qy 121 TCTCAGTTCAGCAAAGAGGAANCGCGTCGCGATGAAAGAGCGCGTCGAAAGGTGCCATCCAG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 41 SerGlnPheSerLysGluArgValAlaMetLysGluAlaLeuLysGlyAlaIleGln 60

Qy 181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly 80

Qy 241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCACGACCA 280
        |||||||||||||||||||||||||||||||||||||||
Db 81 -AsnThrPheIleLysSerPheLeuGlnTrpSerAlaPro 93

; TITLE OF INVENTION: ISOLATED HUMAN PROTEASE PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PROTEASE PROTEINS, AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CL000425
; CURRENT APPLICATION NUMBER: US/60/195,044
; CURRENT FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 179
; LENGTH: 79
; TYPE: PRT
; ORGANISM: HUMAN
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)..(79)
; OTHER INFORMATION: Xaa = Any Amino Acid

US-60-195-044-179

Alignment Scores:
Pred. No.:   2.13e-29
Score: 410.00
Percent Similarity: 100.00%   Length: 79
Best Local Similarity: 100.00%   Matches: 77
Query Match: 14.84%   Conservative: 0
DB: Mismatches: 0
                                          Indels: 0
                                          Gaps: 0

US-10-014-896-1 (1-1509) x US-60-195-044-179 (1-79)

Qy 259 TTTCCTACAGTGGTCACCACCAGCTTTATCCAGCATGAAGCTGCGGAAGAGTATAGCCAC 318
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2 PheProThrValValSerThrSerPheIleGlnHisGluValAlaValGluLeuTyrSerHis 21

Qy 319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTT 378
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 22 LeuPheThrIleGlnGlySerAspProSerLeuGlnProTyrLeuLeuMetAlaHisPhe 41

Qy 379 GATGTGGTGCCCTGCCCCTGAAGAAGCCTGGGAGGTGCCCCCATTCTCTGGGTTGGGAGCGT 438
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 42 AspValValProAlaProGluGluGlyTrpGluValProProProPheSerGlyLeuGluArg 61

Qy 439 GATGGCCGTCATCTATGGTCGGGGGCACACTGGACGACAAGAACTGTGAG 489
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db 62 AspGlyValIleTyrArgGlyThrLeuAspProAspLysAsnSerValMet 78

Search completed: June 27, 2003, 11:18:25
Job time : 376.5 secs

RESULT 15
US-60-195-044-179
; Sequence 179, Application US/60195044
; GENERAL INFORMATION:
; APPLICANT: Bonazzi, Vivien

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 27, 2003, 11:02:29 ; Search time 160.5 Seconds
                                             (without alignments)
                                             4878.737 Million cell updates/sec

OM nucleic - protein search, using frame_plus_n2p model

Title:
Perfect score: 2762
Sequence: 1 atggctcaagcggtgcgtttg..........ctcacctgcacaaactgtga 1509

Scoring table: BLOSUM62
                 Xgapop 10.0 , Xgapext   0.5
                 Ygapop 10.0 , Ygapext   0.5
                 Fgapop   6.0 , Fgapext   7.0
                 Delop   6.0 , Delext   7.0

Searched: 1184592 seqs, 259455700 residues

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Total number of hits satisfying chosen parameters: 2369184

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104416_10519/app_query.fasta_1.1671
-DB=Pending_Patents_AA_New -OFMT=fastan -SUFFIX=n2p.rapn -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cd1 -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US10014896 -ECGN_1_1_68 @runat_27062003_104416_10519
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   Pending_Patents_AA_New:*
             1: /cgn2_6/ptodata/1/paa/PCT_NEW_COMB.pep:*
             2: /cgn2_6/ptodata/1/paa/US06_NEW_COMB.pep:*
             3: /cgn2_6/ptodata/1/paa/US07_NEW_COMB.pep:*
             4: /cgn2_6/ptodata/1/paa/US08_NEW_COMB.pep:*
             5: /cgn2_6/ptodata/1/paa/US09_NEW_COMB.pep:*
             6: /cgn2_6/ptodata/1/paa/US10_NEW_COMB.pep:*
             7: /cgn2_6/ptodata/1/paa/US60_NEW_COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	2322	84.1	473	6	US-10-275-107-69	Sequence 69, Appl
2	1824	66.0	361	1	US-10-219-382A-1141	Sequence 1141, Ap
3	1816	66.0	361	6	PCT-US02-32850A-22	Sequence 22, Appl
4	1816	65.7	361	6	US-10-094-749-2952	Sequence 2952, Ap
5	1796	65.7	361	6	US-10-295-027-201	Sequence 201, App
6	827	65.9	361	7	US-60-452-680-15076	Sequence 15076, A
7	827	29.9	166	6	US-10-218-140-218	Sequence 218, App
8	562.5	20.4	120	6	US-10-219-382-2481	Sequence 2481, Ap
9	562.5	20.4	120	6	US-10-219-382A-2481	Sequence 2481, Ap
10	340.5	12.3	440	6	US-10-156-761-9438	Sequence 9438, Ap
11	318	11.5	441	6	US-10-156-761-14162	Sequence 14162, A

US-10-014-896-1 (1-1509) x US-10-275-107-69 (1-473)

ALIGNMENTS

RESULT 1
US-10-275-107-69
Sequence 69, Application US/10275107
GENERAL INFORMATION:
APPLICANT: PLOWMAN, GREGORY D.
APPLICANT: WHITE, DAVID
APPLICANT: SUDARSANAM, SUCHA
APPLICANT: MANNING, GERARD
APPLICANT: CAENEPEEL, SEAN R.
APPLICANT: PAYNE, VILLA
TITLE OF INVENTION: NOVEL PROTEASES
FILE REFERENCE: 038602/1479
CURRENT APPLICATION NUMBER: US/10/275,107
CURRENT FILING DATE: 2003-11-03
PRIOR APPLICATION NUMBER: PCT/US01/14431
PRIOR FILING DATE: 2001-05-04
PRIOR APPLICATION NUMBER: 607201,879
PRIOR FILING DATE: 2000-05-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 69
LENGTH: 473
TYPE: PRT
ORGANISM: Homo sapiens
US-10-275-107-69

Alignment Scores:
Pred. No.:   1.55e-164
Score: 2322.00
Percent Similarity: 93.45%
Best Local Similarity: 93.25%
Query Match: 84.07%

Length: 473
Matches: 470
Conservative: 1
Mismatches: 2
Indels: 31
Gaps: 2

12	297	10.8	112	6	US-10-219-382-1141	Sequence 1141, Ap
13	297	10.8	112	6	US-10-219-382A-1141	Sequence 1141, Ap
14	274	9.9	422	6	US-10-454-437-202	Sequence 202, App
15	266.5	9.6	458	1	PCT-US02-32727-2506	Sequence 2506, Ap
16	266.5	9.6	458	6	US-09-978-825-2506	Sequence 2506, Ap
17	266.5	9.6	458	6	US-10-057-498-2506	Sequence 2506, Ap
18	226.5	8.2	446	6	US-10-424-599-239368	Sequence 239368, Ap
19	211	7.6	376	1	PCT-US02-36123-1620	Sequence 1620, Ap
20	210	7.6	467	6	US-10-424-599-211620	Sequence 211620, Ap
21	209.5	7.6	474	6	US-10-425-114-47811	Sequence 47811, A
22	209	7.6	474	6	US-10-425-114-72430	Sequence 72430, A
23	209	7.6	461	6	US-10-425-114-42127	Sequence 42127, A
24	208.5	7.6	463	6	US-10-369-493-7208	Sequence 7208, Ap
25	207	7.5	474	6	US-10-425-114-62643	Sequence 62643, A
26	201	7.3	362	6	US-10-369-493-143921	Sequence 143921, A
27	201	7.3	414	1	PCT-US02-36123-5366	Sequence 5366, Ap
28	201	7.3	418	6	US-10-092-411A-3552	Sequence 3552, Ap
29	199.5	7.2	412	6	US-10-950-084-3959	Sequence 3959, Ap
30	199	7.2	410	6	US-10-369-493-21465	Sequence 21465, A
31	198.5	7.2	439	6	US-10-264-213-221	Sequence 221, APP
32	198.5	7.2	446	6	US-10-424-599-211102	Sequence 211102, A
33	195.5	7.1	406	6	US-10-431-652-5309	Sequence 5309, Ap
34	195	7.1	391	6	PCT-US02-36123-2902	Sequence 2902, Ap
35	195	7.1	460	1	PCT-US02-36123-2902	Sequence 2902, Ap
36	194	7.0	388	6	US-10-335-977-7594	Sequence 7594, Ap
37	194	7.0	425	5	US-09-134-000C-3861	Sequence 3861, Ap
38	194	7.0	425	6	US-09-134-000C-3861	Sequence 3861, Ap
39	194	7.0	425	6	US-10-434-665-3861	Sequence 3861, Ap
40	193	7.0	632	6	US-10-366-683-17148	Sequence 17148, A
41	193	7.0	632	6	US-10-419-128-17148	Sequence 17148, A
42	193	7.0	373	1	PCT-US03-02500-52	Sequence 52, Appl
43	190	6.9	388	6	US-10-447-013-4	Sequence 4, Appl
44	190	6.9	388	6	US-10-447-013-4	Sequence 4, Appl
45	189	6.8	394	6	US-10-375-039-44	Sequence 44, Appl

```
Qy      1 ATGGCTCAGCGCGTGCGTGGCGTGCTGGGCCTCCGGTATGCTGTGCTCCTAGTTTTCCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaGlnArgCysValCysValLeuAlaLeuAlaMetLeuLeuLeuValPhePro 20

Qy     61 ACCGTCTCCAAATCCAATGGCCCCAGGACCGGGGAGCAGCAACAAAGGGGGTCCGGAATCCT 120
          ||||:::|||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ThrValSerArgSerMetGlyProArgSerGlyGluTyrGlnArgAlaSerArgIlePro 40

Qy    121 TCTCAGTTCAGCAAAGAGAACAGCGCTCGGCAGTCGGGATGAAAGAGGGTGGCCATCCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerGlnPheSerLysGluGlyArgValAlaMetLysGluLysGlyGlyAlaIleGln 60

Qy    181 ATTCCAACAGTGACTTTACCTCTCAGAAGTCCAATACTACACGCCCTGGCTCAGTTCGGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly 80

Qy    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
          ||||||||||||     |||||||||||||||||||||||||||||||||||||||||||
Db     81 LysTyrIleArgLysValPheProThrValValSerThrSerPheIleGlnHisGluVal 100

Qy    301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCCTCGGGACCCCCAGCTTGCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 ValGluGluTyrSerHisLeuPheThrIleGlnGlyLeuGlyThrProSerLeuGlnProTyr 120

Qy    361 CTGCTGATGGCTCACTTTGATGTGGTGCCCCCTGAAGAAGGCTGGGGAGGTGCCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGlyValProPro 140

Qy    421 TTCTCTGGGTTGGGAGGCTGAIGGCGTCATCTCTATGGTCTCGGGGCCACACTGGACGACAAGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 PheSerGlyLeuGluArggAspGlyValIleTyrGlyValGlyThrLeuLeuAspaspLysAsn 160

Qy    481 TCTGTGATGGCATTACTGCAGGCCTTGGGAGCTCCCTGCTGATCAGGAAGTACATCCCCGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 SerValMetAlaLeuLeuGlnAlaLeuGluLeuLeuArgGlyAlaSerTyrIleProArg 180

Qy    541 AGAtCTTTCTTCAtTTCTCTGGGCCAATGATGAGGAGTCATCAGGGACAGGGGGTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ArgSerPhePheIleSerLeuGlyHisAspGluGluSerSerGlyThrGlyAlaGlnArg 200

Qy    601 ATCTCAGCCCTGCTACAGTCAAGGGGCGGTCCAGCTTCATTGTGGACGAGGGGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 IleSerAlaLeuLeuGlnSerArgGlyValGlnAlaPheIleValAspGluGlyGly 220

Qy    661 TTCAtCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCAtCGGCCTTGAtTCGCAGTCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 PheIleLeuAspAspPheIleProAsnPheLysLysProIleAlaLeuIleLeuAlaValSer 240

Qy    721 GAGAGGGTTCCATGAACCTCAtGGTCGCAGTAAAACATGACTTCAGGGCCCACTCTTCAGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla 260

Qy    781 CCTCCAAAGGAGGACAAGCAtTCGGCAtCCTTGCCAGCTGCTGTCACCCGAGtTGGACAGACA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ProProLysGluThrSerIleGlyIleLeuAlaAlaAlaValSerArgLeuGluGlnThr 280

Qy    841 CCAATGCCTAtATAtTTGGAAGGCGGGACAGtGTGGGtGACtGTATGTGCACGAAtTGGGCAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ProMetProIleIlePheGlySerGlyThrValValThrValCysAlaArgLeuGlyGln 300

Qy    901 GAGtTTCCCTTCCCTGTGCAAATATAATCCTGAGCAACCCAtGGCtAtTtGAACCACtTATA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    300 ---------------------------------------------------------- 300

Qy    961 AGCAGGGTTTAtGGAGGAGAAAACCCTtAACCAAtGCAAtAAtCAGGACCACACGGCGCACTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||
Db    301 Glu-ValTyrGlyLysSerLeuAsnGlnCysAsnAsnGlnAspHisHisGlyThrHr 320

Qy   1021 ACCAtAtTtCAAACGCAGGGtCAAGTtCAAtGtCAtCCCCAGtTGGCCCCAGtTGGCCCACAGtC 1080
          ||||||||||||||||||   ||
Db    320 IsHisIleGlnSerArg-Val-------------------------------------- 331
```

```
Qy   1081 AACTTCCGGAtTCACCCTGGACAGACAGtCCAAGAGGGtCCTAGtGAACTCACGAAGACAtT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    332 AsnPheArgIleHisProGlyGlnThrArgValGlnGluValLeuGluLeuThrLysAsnIle 351

Qy   1141 GTGGCTGAtAACAGAGtCCAGtTCCAtGtGttCAAtGtGTtGAGtGCCtTtGACCCCTCCCGtCAGC 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    352 ValAlaAspAsnArgValGlnPheHisValLeuAsnSerAlaPheAspProLeuProValSer 371

Qy   1201 CCtTCtGAtGACAAGGCCtTGGGtCtACCAGCtGtCtCCCCCGACAtGttGACtGtACAGtCCGTCTTC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    372 ProSerAspAspLysAlaLeuGluTyrGlnLeuLeuArgGlnThrValGlnSerValPhe 391

Qy   1261 CCGGAAGtCAAtAtTACtGCCCCAGtTACtTCTAtTtGGGAACACACAGACAGACGCCGAtTtCtT 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    392 ProGluValAsnIleThrAlaProValThrSerIleGlyAsnThrAspSerArgArgPhePhe 411

Qy   1321 ACAAAtCTtCCACCACtGGtCAtCCTACtACtGGtTtCTtACCCCCACtCtACAtAtCACtGAAtGACTtC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    412 ThrAsnLeuThrThrGlyIleLeuTyrArgProPheTyrProIleLeuGlnProGluAspPhe 431

Qy   1381 AAACGACAtCCAtGGAGtCAAGCAGAGAAAAtCtCAGtCtCAAGCCtAtTGAGACCCCAAGtGAAA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    432 LysArgIleHisGlyValAlaSnGluLysIleSerSerValGlnAlaIleGluThrGlnLys 451

Qy   1441 TtCAtCtTtGAGtTtGAtTCAGAAtGCTGtGACACAGCACCAGGAGGAGCAGtTtCtCtCAtCCTGCAC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    452 PheIlePheGluLeuIleGlnAsnAlaAspProHisArgAspGlyGluProValSerHisSerLeuHis 471

Qy   1501 AAACTG 1506
          ||||||
Db    472 LysLeu 473
```

RESULT 2
PCT-US02-32850-22
; Sequence 22, Application PC/TUS0232850
; GENERAL INFORMATION:
; INCYTE GENOMICS, INC.
; APPLICANT: RAMKUMAR, Jayalaxmi
; APPLICANT: GORVAD, Ann E.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: EMERLING, Brooke M.
; APPLICANT: YANG, Junming
; APPLICANT: LEE, Soo Yeun
; APPLICANT: TRAN, Uyen K.
; APPLICANT: BECHA, Shanya D.
; APPLICANT: DUGGAN, Brendan M.
; APPLICANT: LEE, Ernestine A.
; APPLICANT: GRIFFIN, Jennifer A.
; APPLICANT: LI, Joana X.
; APPLICANT: SPRAGUE, William W.
; APPLICANT: HAFALIA, April J.A.
; APPLICANT: CHAWLA, Narinder K.
; APPLICANT: LEHR-MASON, Patricia M.
; APPLICANT: KABLE, Amy E.
; APPLICANT: YUE, Henry
; APPLICANT: MARQUIS Joseph P.
; APPLICANT: YAO, Monique G.
; APPLICANT: RICHARDSON, Thomas W.
; APPLICANT: TANG, Y. Tom
; APPLICANT: JIN, Pei
; APPLICANT: CHIEN, David
; APPLICANT: BHATIA, Umesh G.
; APPLICANT: BURRILL, John D.
; APPLICANT: LEE, Sally
; APPLICANT: BLAKE, Julie J.
; APPLICANT: HO, Anne
; APPLICANT: ZHENG, Wenjin
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE PROTEINS
; FILE REFERENCE: PF-1237 PCT
; CURRENT APPLICATION NUMBER: PCT/US02/32850
; CURRENT FILING DATE: 2002-10-11
; PRIOR APPLICATION NUMBER: US 60/329,689
; PRIOR FILING DATE: 2001-10-12

```
; PRIOR APPLICATION NUMBER: US 60/335,703
; PRIOR FILING DATE: 2001-10-25
; PRIOR APPLICATION NUMBER: US 60/348,887
; PRIOR FILING DATE: 2001-11-09
; PRIOR APPLICATION NUMBER: US 60/334,145
; PRIOR FILING DATE: 2001-11-28
; PRIOR APPLICATION NUMBER: US 60/337,451
; PRIOR FILING DATE: 2001-12-06
; PRIOR APPLICATION NUMBER: US 60/340,584
; PRIOR FILING DATE: 2001-12-14
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PERL Program
;
; SEQ ID NO 22
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 3118830CD1

PCT-US02-32850-22

Alignment Scores:
  Pred. No.:
  Score:                  1824.00
  Percent Similarity:     93.998
  Best Local Similarity:  93.998
  Query Match:            66.04%
  DB:                     1

US-10-014-896-1 (1-1509) x PCT-US02-32850-22 (1-361)
```

```
Qy      1 TCTCAGTTCAGCAAAAGAAGAGGAACGACGTCCGCGATGAAAGAGGCGCTGAAAGGTTGCCATCCAG  60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTCAGCGCGGTGAGGGTTGCGTTTGGCTGGCCCTTGGCTATGGCTGCTCCTCCTTAGTTTTTCCCT  60

Qy     21 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro  40
Db      1 MetAlaGlnArgCysValCysValCysValLeuAlaLeuValAlaMetLeuLeuLeuValPhePro  20

Qy     61 ACCGTCTCCAGATCGATGGGCCCGAGGAGGGGGACGCATCAAAAGGCGTCGCGAATCCCT  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ...

Qy    121 ATTCCAACAGTGACTTTTAGCTCTTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA  180
Db     41 SerGlnPheSerLysGluGluArgValAlaMetLysGluValGlyAlaLeuGln  60

Qy    181 ATTCCAACAGTGACTTTTAGCTCTTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA  240
Db     61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly  80

Qy    241 AAATACATTCATAAAGTCTTTCCTACACGTGGTTCAGCACCACAGCTTTATCCAGCATGAAGTC  300
Db     81 LysTyrIleHisLysValPheProThrValValSerThrSerPheIleGlnHisGluVal  100

Qy    301 GTGGAAGAGTATAGCACCACCTGTTCACTATCCAAGGCTCGGAACGCTTGCAGCCCTAC  360
Db    101 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuProTyr  120

Qy    361 CTGCTCGATGGCTCACTTTGATGTGGTGCGTGCCTGCCCCTGAAGAAGGCTGGGAGGTGGCCCCCA  420
Db    121 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro  140

Qy    421 TTCTCTGGGTTGGAGCGGTCGGGCGTCATCCTATGGGTCGGGGGCACACTGCACGACAGAAC  480
Db    141 PheSerGlyLeuGluArgAspGlyValIleLeuTyrGlyArgGlyThrLeuAspAspLysAsn  160

Qy    481 TCTGTGATGGCAATTACTACTGCAGGCCTTGGAGCTCCTGCTGATCAGAAGTACATCCCCCCGA  540
Db    161 SerValMetAlaLeuLeuLeuGlnAlaLeuGluLeuLeuLeuArgLysTyrIleProArg  180

Qy    541 AGATCTTTCTTCATTTCTCTGGGCCATGAATGAGGAGACCAGGGGGTTCAGAGG  600
Db    181 ArgSerPhePheIleSerLeuGlyHisAspGluGlnSerSerGlyThrGlyAlaGlnArg  200

Qy    601 ATCTCAGCCCTGCTACAGTCAAAGGGGGCGGTTCCAGTTGACGAGGGGGGC  660
```

Right column:

```
Db    201 IleSerAlaLeuLeuGlnSerArgGlyValGlnLeuUalaPheIleValAlaAspGluGlyGly  220

Qy    661 TTCACTCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATTCGCCTTGATTTGCAGTCTCA  720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 PheIleLeuUaaspAspPheIleProAsnPheLysLysProIleAlaLeuUaIleUalSer  240

Qy    721 GAGAAGGGTTCCATGAACCTCATGGTGCAAGTAAACATGACTTCAGGCCGCACTCTTCAGCT  780
Db    241 GluLysGlySerMetAsnLeuMetUalGlnUalAsnMetThrSerGlyArgHisSerSerAla  260

Qy    781 CCTCCAAAAGGAGACAACAGCCATTGGCANTCCTTCGCAGCTGCTCTTCAGCCGACTTGGAGCAGACA  840
Db    261 ProProLysGluThrSerIleGlyIleLeuUaIleUalAlaSerArgLeuGluGlnThr  280

Qy    841 CCAATGCCTACATCATATTTGGAAGCGGGACAGCTGGTGGTGGACTGTATTGCAGCAACTGGCAAAT  900
Db    281 ProMetProIleIlePheGlySerGlyThrUaIleUalUalUaAspCysIleAlaAlaThrGlyLysAsn  300

Qy    901 GAGTTTCCCTTCCCTGTCAATAAAATCCTGAGCAACACCATGGCTATTTGAACCACTTATA  960
Db    301 Glu--UalTyrGlyUalLysSerLeuAsnGlnCysAsnAsnGlnAsnProUalValThrHisH  320

Qy   1021 ACCATATTCAAAACAGGGGGTCAAGTTCATCCCCCCAGGTGCCCAAGGCCACAGTC  1080
Db    320 lsHisSerArgGlyUalnalGlnCysHisProProSerGlyProGlyHisSerG  340

Qy   1081 AACTTCCGGATTCACCCGGACAGACAGTCCAAGAGGTCCTAGAAACTCACCAGAAGAACATT  1140
Db    340 lnLeuProAspSerProTrpThrAspSerProArgGlyLyProArgThrHisGluHisSer  360

Qy   1141 GTGGCC  1145
          ||||
Db    360 ysGly  361
```

```
RESULT 3
PCT-US02-32850A-22
Sequence 22, Application PC/TUS0232850A

GENERAL INFORMATION:
APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: GORVAD, Ann E.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: EMERLING, Brooke M.
APPLICANT: YANG, Junming
APPLICANT: LEE, Soo Yeun
APPLICANT: TRAN, Uyen K.
APPLICANT: BECHA, Shanya D.
APPLICANT: DUGGAN, Brendan M.
APPLICANT: LEE, Ernestine A.
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: LI, Joana X.
APPLICANT: SPRAGUE, William W.
APPLICANT: HAFALIA, April J.A.
APPLICANT: CHAWLA, Narinder K.
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: KABLE, Amy E.
APPLICANT: YUE, Henry
APPLICANT: MARQUIS Joseph P.
APPLICANT: YAO, Monique G.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: TANG, Y. Tom
APPLICANT: JIN, Pei
APPLICANT: CHIEN, David
APPLICANT: BHATIA, Unesh G.
APPLICANT: BURRILL, John D.
APPLICANT: LEE, Sally
APPLICANT: BLAKE, Julie J.
APPLICANT: HO, Anne
APPLICANT: ZHENG, Wenjin
```

TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
FILE REFERENCE: PF-1237 PCT
CURRENT APPLICATION NUMBER: PCT/US02/32850A
CURRENT FILING DATE: 2002-10-11
PRIOR APPLICATION NUMBER: US 60/329,689
PRIOR FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/335,703
PRIOR FILING DATE: 2001-10-25
PRIOR APPLICATION NUMBER: US 60/348,887
PRIOR FILING DATE: 2001-11-09
PRIOR APPLICATION NUMBER: US 60/334,145
PRIOR FILING DATE: 2001-11-28
PRIOR APPLICATION NUMBER: US 60/337,451
PRIOR FILING DATE: 2001-12-06
PRIOR APPLICATION NUMBER: US 60/340,584
PRIOR FILING DATE: 2001-12-14
NUMBER OF SEQ ID NOS: 80
SOFTWARE: PERL Program

SEQ ID NO 22
LENGTH: 361
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 3118830CD1

PCT-US02-32850A-22

Alignment Scores:

Pred. No.:
Score: 1824.00
Percent Similarity: 93.99%
Best Local Similarity: 93.99%
Query Match: 66.04%
DB: 1

Length: 361
Matches: 360
Conservative: 0
Mismatches: 1
Indels: 22
Gaps: 1

US-10-014-896-1 (1-1509) x PCT-US02-32850A-22 (1-361)

RESULT 4
US-10-094-749-2952
Sequence 2952, Application US/1009474
GENERAL INFORMATION:
APPLICANT: ISOGAI, TAKAO
APPLICANT: SUGIYAMA, TOMOYASU
APPLICANT: OTSUKI, TETSUJI
APPLICANT: WAKAMATSU, AI
APPLICANT: SATO, HIROYUKI
APPLICANT: ISHII, SHIZUKO
APPLICANT: YAMAMOTO, JUN-ICHI
APPLICANT: ISONO, YUUKO
APPLICANT: HIO, YURI
APPLICANT: OTSUKA, KAORU
APPLICANT: NAGAI, KEIICHI
APPLICANT: IRIE, RYOTARO
APPLICANT: TAMECHIKA, ICHIRO
APPLICANT: SEKI, NAOHIKO
APPLICANT: YOSHIKAWA, TSUTOMU
APPLICANT: OTSUKA, MOTOYUKI
APPLICANT: NAGAHARI, KENJI
APPLICANT: MASUHO, YASUHIKO
TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
FILE REFERENCE: 084335/0160
CURRENT APPLICATION NUMBER: US/10/094,749
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 60/350,435
PRIOR FILING DATE: 2002-01-24
PRIOR APPLICATION NUMBER: JP 2001-328381

; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 2952
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-094-749-2952

Alignment Scores:
Pred. No.:
Score: 1816.00
Percent Similarity: 93.73%
Best Local Similarity: 93.73%
Query Match: 65.75%

Length: 361
Matches: 359
Conservative: 0
Mismatches: 2
Indels: 22
Gaps: 1

US-10-014-896-1 (1-1509) x US-10-094-749-2952 (1-361)

Qy   1 ATGGCTCAGCGGGGTGGCCGTTGCGGTGGCTATGCCTCCTAGTTTTCCCT 60
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaGlnArgGlyGlyValCysValLeuAlaLeuLeuLeuLeuValPhePro 20

Qy   61 ACCGTCTCCAGATCAGATCGGGCCCGAGGAGCGCGGGGAGGCATCAGAAATCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 40

Qy 121 TCTCAGTTCAGCAAAAGAGGAACGGCGTCGCGATGAAAGAGAGGTGCCATCCAG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerGlnPheSerIleSerGluGlyArgGlyValAlaMetLysGluLysGlyAlaIleGln 60

Qy 181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAATACTACAGCCCTGGCTTCAGTTCGGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IleProThrValThrPheSerSerGluLysSerLysTyrTyrSerProGlyPheGly 80

Qy 241 AAATACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 LysTyrIleHisLysValPheProThrValSerThrSerPheIleGlnHisGluVal 100

Qy 301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAAGCCTGGACCCCAGCTTGAGCCCTAC 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 101 ValGluGluTyrSerHisLeuPheThrIleGlnSerAspProSerLeuSerProTyr 120

Qy 361 CTGCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGCGCTGGGAGTGCCCCCA 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 LeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro 140

Qy 421 TTTCTCTGGGTTGGAGCGTGATGCGGCGTCATCTATGGTCGGGGACACTGGACAAGAAC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 141 PheSerGlyLeuGluArgAspGlyValIleTyrGlyTyrGlyThrLeuAspAspLysAsn 160

Qy 481 TCTGTGATGGCAATACTGGAGCCTCCTGCCTGATCAGGAAGTACATCCCCGA 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 161 SerValMetAlaLeuLeuGlnAlaLeuGluLeuLeuLeuArgLysTyrIleProArg 180

Qy 541 AGATCTTTCTTCATTTCTCTGGGCCCATGATGAGGAGACAGGGGCTCAGAGG 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ArgSerPhePheIleSerLeuGlyHisAspGluGluSerGlyTyrGlyAlaGlnArg 200

Qy 601 ATCTCAGCCCCTGCTCTACAGTGTCAAGGGGGGTCCAGGTTACTTCATTGTGGACGAGGGGGC 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 201 IleSerAlaLeuLeuGlnSerArgGlyValGlnLeuAlaPheIleValAspAspGlyGly 220

Qy 661 TTCACTCTTGGATGATTCAATCCTCAAACTTCAAAGGAAGCCCATCGCCTTGATTGCAGTCTCA 720
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 221 PheLeuLeuAspAspPhePheIleProAsnPheLysLysLysProIleAlaLeuIleAlaValSer 240

Qy 721 GAGAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCT 780
        ||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla 260

Qy 781 CCTCCAAAGAGACAAGACATTGGCATCCTTGCAGCTGCTGTCGTGGAGCAGACA 840
        ||||||||||||||||||||||||||||||||||||||||||||||||||

Db 261 ProProLysGluThrSerIleGlyTyrIleLeuAlaAlaAlaValSerArgLeuGluGlnThr 280
Qy 841 CCAATGCCTATCATATATTGGAAGCGGGGACAAGTGGTGACGTGATTGCAGCAAACTGGCAAAT 900
Db 281 ProMetProIleIlePheGlySerGlyThrValValThrValLeuGlnLeuLeuAlaAsn 300
Qy 901 GAGTTTCCCTTCCCTGTCAATPATATAACCTGAGCAACCATGGCTATTTGAACCACTTATA 960
Db 300 -- 300

Qy 961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATATCAAAATAATATCAGGACCACCACGCACTC 1020
        |||||||||||||||||||||||||||||||||||||||||||||||
Db 301 Glu--ValTyrGlyGluLysSerLeuAsnGlnAsnAsnGlnAsnAspHisHisGlyTyrThrH 320

Qy 1021 ACCATATTCAAAGCAGGGGGTCAAGTTCAAATGCATCCCCCCCAGTGCCCAGGCCACAGTC 1080
        ||||||||||||||||||||||||||||||||||||||||||||||
Db 320 IleHisIleGluGlnSerArgGlyValGlnValGlyHisSerProSerGlyProGlyHisSerG 340

Qy 1081 AACTTCCGGATTCACCCTGGACAACAGCAGTCCAAGAGGAGGTCCTAGAACTCACGAAGAGAACATT 1140
        |||||||||||||||||||||||||||||||||||||||||||||
Db 340 IleLeuProAspSerProTrpHisArgSerProArgGlyProArgThrHisGluGluHisC 360

Qy 1141 GTGGC 1145
        |||||
Db 360 ySGly 361

RESULT 5
US-10-295-027-201
; Sequence 201, Application US/10295027
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-0125000US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 201
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-201

Alignment Scores:

Pred. No.:	7.9e-127
Score:	1816.00
Percent Similarity:	93.73%
Best Local Similarity:	93.73%

Length:	361
Matches:	359
Conservative:	0
Mismatches:	2
Indels:	2
Gaps:	1

Query Match: 65.75%
DB:

US-10-014-896-1 (1-1509) x US-10-295-027-201 (1-361)

```
Qy      1 ATGGCTCAGCGGTGGGCGTTTGCGTTGCCTGGCTATGCTGCTCCTTAGTTTTCCCT   60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaGlnArgCysValCysValLeuAlaLeuValAlaMetLeuLeuLeuValPhePro   20

Qy     61 ACCGTCTCCAGATCGATGGGCCCGGAGGAGGCCATCAAAGGCGTCGCGAATCCCT  120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 ThrValSerArgSerMetGlyProGluGluAlaIleLysGlyValAlaSerArgIlePro   40

Qy    121 TCTCAGTTCAGCAAAGAGGAAACGCGTCGCGATGAAAAGGACGTGCCATCCAG  180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerGlnPheSerLysGluGluArgValAlaMetLysGluValAlaIleGln   60

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATAACTACAGCCCCTGGGTTCGGA  240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 IleProThrValThrPheSerSerGluLysSerAsnAsnTyrSerProTrpValArgGly   80

Qy    241 AAATACATTCATAAAGTCTTTTCCTACAGTGGTCAGCACCACAGCTTTATCCAGCATGAAGTC  300
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 LysTyrIleHisLysValPheSerTyrSerGlyGlnHisHisSerPheIleGlnHisGluVal  100

Qy    301 GTGGAAGAGTATAGCCACCCTGTTCACTATATCCAAAGGCGTCGGACCCCAGCTTGCAGCCCTAC  360
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 ValGluGluTyrSerHisProValHisTyrIleGlnGlySerAspProSerLeuGlnProTyr  120

Qy    361 CTGCTGATGGCTCACTTTGATGTTGGTGCCTGCCCTGAAGAAAGGCTGGGAGGTGCCCCA  420
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 LeuLeuMetAlaHisPheAspValGlyAlaProGluGluArgLeuGlyTrpGluValProPro  140

Qy    421 TTCTCTTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGGCGCACACTGGACGACAAGAAC  480
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    141 PheSerGlyLeuGluArgAspGlyValIleTyrGlyTyrGlyThrHisTrpLeuAspAspLysAsn  160

Qy    481 TCTGTGATAGGCAATTACTGCAGGCCTTGGAGTCCTCCTGCTCATCAGGAAGTACATCCCCGA  540
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    161 SerValMetIleAlaLeuLeuGlnLeuLeuLeuArgLysTyrIleProArg  180

Qy    541 AGATCTTTCATTTCTCTGGGCCATGATGAGGAGTGACAGGGGGCTCAGAGG  600
            |||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ArgSerPhePheIleSerLeuGlyHisAspGluGluSerGlyThrGlyAlaGlnArg  200

Qy    601 ATCTCAGCCCCTGCTCACAGTCAAAGGGGGCGTCCAGCTACCTTCATTGTGGACGAGGGGGC  660
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    201 IleSerAlaLeuLeuHisGlnSerArgGlyValGlnLeuProPheIleValAspGluGlyGly  220

Qy    661 TTCATCTTGGATAGGATTTCAATTCCTAACTTTCAGGAAGGCCCATACGCGCCTTGATTGCAGTCTCA  720
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    221 PheIleLeuAspArgPheIleProAsnPheLysLysGlyProTyrAlaLeuAspCysSer  240

Qy    721 GAGAAGGGTTCCAATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCT  780
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla  260

Qy    781 CCTCCAAAGGAGACAAAGCAATTGGCATCCTTGCAGCTCCTGCTGCCGGATTGGAGAGACA  840
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    261 ProProLysGluThrLysGlnLeuAlaSerLeuGlnLeuLeuLeuProAspTrpArgAspThr  280

Qy    841 CCAATGCCTATCATAATTTGGAACGCGGACAGTCGTGACTGTTATTGCAGCAACTGGCAAAT  900
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    281 ProMetProIleIlePheGlySerGlyThrValThrValLeuGlnLeuLeuAlaAsn  300

Qy    901 GAGTTCCCTCCCTGTCAAATATAATCCTGAGCAACCCATGGCTATTGAACCACTTATA  960
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GAGTTCCCTCCCTGTCAAATATAATCCTGAGCAACCCATGGCTATTGAACCACTTATA  960
```

```
Db    300 ----------------------------------------------------------  300
```

```
Qy    961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATATATCAGGAGACCACCACGGCACTC  1020
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 Glu--ValTyrGlyGluLysSerLeuAsnGlnCysAsnAsnGlnAsnHisHisGlyThrH  320

Qy   1021 ACCATATTCAAAGCACGGGGGTCCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCGCCACATC  1080
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    320 IleHisIleGlnSerArgGlyValGlnValGlnCysHisIleProProProSerGlyProSerGlyTyrProGlyHisSerG  340

Qy   1081 AACTTCCGGGATTCACCCCTGGACAGAGCAGAGCAGTGCTCAAGAGGTCCTAGAACTACAGGAGAACAATT  1140
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    340 InLeuProCysAspSerProTrpThrAspSerProArgGlyProGlyArgGlyThrHisSerHisGlyGluHisSC  360

Qy   1141 GTGGC  1145
            |||||
Db    360 ysGly  361
```

RESULT 6
US-60-452-680-15076

; Sequence 15076, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele
; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 15076
; LENGTH: 361
; TYPE: PRT
; ORGANISM: Homo sapiens
US-60-452-680-15076

Alignment Scores:

Pred. No.:	2.45e-125
Score:	1796.00
Percent Similarity:	99.45%
Best Local Similarity:	99.45%

Length:	361
Matches:	359
Conservative:	0
Mismatches:	2
Indels:	1
Gaps:	0

Query Match: 65.03%
DB:

US-10-014-896-1 (1-1509) x US-60-452-680-15076 (1-361)

```
Qy    194 CTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCCT-GAGTTCGGAGAAAATACATTCAT  252
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 LeuLeuAlaLeuArgSerProIleLeuGlnProTrpPAlaGluPheGlyLysTyrIleHis  20

Qy    253 AAAGTCTTTCCTACAGTGGTCAGCACCACAGCTTTATCCAGCATGAAGTCGTGGAAGAGTAT  312
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     21 LysValPheProThrValSerThrSerProPheIleGlnHisGluValValGluGluTyr  40

Qy    313 AGCCACCCTGTTCACTATATCCAAAGGCTCGGACCCCAGCTTGCAGCCCTACCCTGCTGATGGCT  372
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     41 SerHisProThrIleGlnGlySerAspProSerLeuGlnProTyrLeuLeuMetAla  60

Qy    373 CACTTTGATGTGGTGCCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTG  432
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 HisPheAspValValProAlaProGluGluGlyTrpGluValProProPheSerGlyLeu  80

Qy    433 GACGGTGAGTGGGGTCATCTATGGTCGGGGCACACTGGACGACACAGAAACTCTGTGATAGGCA  492
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     81 GluArgAspGlyValIleTyrGlyArgGlyThrLeuAspAspProLysAsnServalMetAla  100

Qy    493 TTACTGCAGGCCTTGGAGCTCCTGCTCATCAGGAAGTACATCCCCGAAGATCTTTCTTC  552
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    101 LeuLeuGlnLeuLeuLeuLeuLeuLeuArgLysTyrIleProArgArgSerPhePhe  120

Qy    553 ATTTCTCTGGGCCCATGATGAGGAGTCATCAGGGGACAGAGGGGCTCAGAGGATCCAGCCCTG  612
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (1)..(1)
; OTHER INFORMATION: "Xaa" = "Any Amino Acid"
US-10-218-140-218

Alignment Scores:
Pred. No.:
Score: 3.81e-53   Length: 166
Percent Similarity: 827.00   Matches: 164
Best Local Similarity: 99.39%   Conservative: 0
Query Match: 99.39%   Mismatches: 1
DB: 29.94%   Indels: 0
                        6   Gaps: 0

US-10-014-896-1 (1-1509) x US-10-218-140-218 (1-166)

Qy   142 CGGGTCCGGATGAAAAGAGGCGGCTGAAAGGTGCCATCCAGAATTCCAACACAGTGACTTTTAGC 201
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2   ArgValAlaMetLysGluAlaGlyGlyLeuLysGlyAlaIleGlnAsnSerAsnThrValThrPheSer 21

Qy   202 TCTGAGAGAGTCCAATACTACACAGCCCTGGCTGAGTTGGGAAAATACATTCATTAAAGTCTTT 261
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   22   SerGluLysSerAsnThrThrAlaLeuAlaGluPheGlyLysTyrIleHisLysValPhe 41

Qy   262 CCTACAGTGGTCAGCACCACCAGCTTTATCCAGGATGAAGTCGTGGAAGAGTATAGCCACCTG 321
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   42   ProThrValValSerThrSerProPheIleGlnAspGluValValGluGluTyrSerHisLeu 61

Qy   322 TTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCTCACCTGCTGATGGCTCACTTTGAT 381
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   62   PheThrIleGlnGlySerAspProSerLeuGlnProHisLeuLeuMetAlaHisPheAsp 81

Qy   382 GTGGTGCCCTGCCCCCTGAACAGAGGCTGGGAGGGTGCCCCCCATTCTCTGGGTTGGAGCTGAT 441
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   82   ValValProAlaProGluGluGlnArgLeuGlyGlyAlaProHisSerLeuGlyLeuGluLeuIle 101

Qy   442 GGCGTCATCTTATGGTGCGGGGCCACACTGGACGACGACAAGAACTCTGTGATGGCATTACTGCAG 501
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 GlyValIleTyrGlyTrpGlyThrLeuAspAspAspLysAsnSerValMetAlaLeuLeuGln 121

Qy   502 GCCTTGGAGCTCCCTGCTCGATCAGCAGAAGTACATCCCCGAAGAGATCTTTTCAATTCTCTG 561
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 AlaLeuGluLeuLeuLeuIleArgLysTyrIleProArgArgSerPhePheSerLeu 141

Qy   562 GGCCATGATGAGGAGTCATCAGGGGACAGCAGGGGCTCAGAGGGATCTCAGCGCCCTGCTACACTCA 621
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 GlyHisAspGluGluSerSerGlyAspSerArgGlyTyrGlyAlaGlnArgIleSerAlaLeuLeuGlnSer 161

Qy   622 AGGGCGGTCCAGCTA 636
          |||||||||||||||
Db   162 ArgGlyValGlnLeu 166

; DB:

; RESULT 7
; US-10-218-140-218
; Sequence 218, Application US/10218140
; GENERAL INFORMATION:
; APPLICANT: Leach, Martin D.
; APPLICANT: Shimkets, Richard A.
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
; TITLE OF INVENTION: ENCODED THEREBY
; FILE REFERENCE: 15966-543 CON
; CURRENT APPLICATION NUMBER: US/10/218,140
; CURRENT FILING DATE: 2002-08-12
; PRIOR APPLICATION NUMBER: 09/540,763
; PRIOR FILING DATE: 2000-03-30
; PRIOR APPLICATION NUMBER: 60/127,728
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/127,636
; PRIOR FILING DATE: 1999-04-02
; PRIOR APPLICATION NUMBER: 60/127,607
; PRIOR FILING DATE: 1999-03-31
; NUMBER OF SEQ ID NOS: 6322
; SOFTWARE: Curanator Version 1.0
; SEQ ID NO 218
; LENGTH: 166

Db   121 IleSerLeuGlyHisAspGluGluSerSerGlyThrGlyAlaGlnAsnArgIleSerAlaLeu 140
Qy   613 CTACAGTCAAGGGGCGGTCCAGCTAGCCCTTCATTGTGACGAGGGGGGCCTTCANCTTGGAT 672
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   141 LeuGlnSerArgGlyValGlnLeuAlaPheIleValAspGluGlyGlyPheIleLeuAsp 160
Db   673 GATTTCATTGTTGACTTCAAGAAGCCCATACGCCCTGATTGACTCTCAGAGAAGGGTTCC 732
Db   161 AspPheIleValAspPheLysLysProIleAlaLeuThrAlaValSerGluLysGlySer 180
Qy   733 ATGAACCTCATGCTGCCAAGTAAACATGACTTCAGGCCACTCTTCAGCTCCTCCAAAGGAG 792
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   181 MetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAlaProProLysGlu 200
Db   793 ACAAGCATTGGCATCCTTCGAGCTGCTGTGCAGCCGGATTGGAGCAGAGCACCAAATGCCTATC 852
Db   201 ThrSerIleGlyIleLeuLeuAlaAlaValSerArgLeuGluGlnThrProMetProIle 220
Qy   853 ATATTTGGAAGCGGGACAGTGGTGAGCTGTGATTGCAGCAACTGCGAAAATGAGTTCCCTTC 912
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   221 IlePheGlySerGlyThrValValThrCysAspCysSerAsnLeuLeuAlaAsnGluPhe 240
Db   913 CCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATAAGCACAGTTTATG 972
Db   241 ProValAsnIleIleLeuSerAsnProTrpLeuPheGluProLeuIleSerArgPheMet 260
Qy   973 GAGAGAAATCCCTTACCAATGCAATAATACAGGACCCACCACCAGCACTCACCACATATTCAAA 1032
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   261 GluArgAsnProLeuThrAsnAlaIleIleGlnAspProThrThrAlaLeuThrThrTyrSerLys 280
Db   1033 GCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGGCCCCAGGCCACAGTCAACTTCCGGATT 1092
Db   281 AlaGlyValLysPheAsnValIleProProValAlaProGlyHisSerGlnLeuProAsp 300
Qy   1093 CACCCTTCGGACACAGACAGCCAAGAGGGTCCTACAACTCACGAGAGACACATTGTGGCTGATAAC 1152
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 HisProGlyHisGlnThrValGlnGlyLeuGlnLeuHisArgAspThrLeuTrpLeuIleThr 320
Db   1153 AGAGTCCAGTTCCAATGTGTTGAGTGGCCTTTGACCCCCTTCAGCCCTTCTGATGAC 1212
Db   321 ArgValGlnPheHisValLeuSerArgLeuaPheAspProLeuProValSerProSerAspAsp 340
Qy   1213 AACGGCCTTGGGGTACCAGCGTGCTCCGGCGACAGCGTACAGTCCGTCTTCCCGGAAGTCAAT 1272
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   341 LysAlaLeuGlyGlyThrTyrGlnLeuLeuLeuArgGlnThrValGlnSerValPheProGluValAsn 360
Db   1273 ATT 1275
          |||
Db   361 Ile 361

; RESULT 8
; US-10-219-382-2481
; Sequence 2481, Application US/10219382
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
; APPLICANT: Ma, Yunging
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted polypeptides
; FILE REFERENCE: 806A
; CURRENT APPLICATION NUMBER: US/10/219,382
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,261
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25

Left column

```
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 2481
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
DB: US-10-219-382-2481

US-10-014-896-1 (1-1509) x US-10-219-382-2481 (1-120)

Alignment Scores:
Pred. No.:
Score:                 1.9e-33        Length:       120
Percent Similarity:    562.50         Matches:      114
Best Local Similarity: 79.72%         Conservative: 0
Query Match:           79.72%         Mismatches:   0
                       20.37%         Indels:       29
                       6              Gaps:         1

Qy    1 ATGGCTCAGCGCGGTGCGTTTTGCGTTTGCTGCTGGCCTATGCTGCTCCTCCTCAGTTTTCCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7 MetAlaGlnArgGlyValCysValLeuLeuLeuLeuValAlaMetLeuLeuLeuValPhePro 26

Qy   61 ACCGTCTCCAGATCGATGGGCCCGAGGAGGAGGCAATCAAAAGGCGTCGCGGAATCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 46

Qy  121 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAAGAGGACGGCTGAAAGGTGCCATCCAG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   47 SerGlnPheSerLysGluGluArgValAlaMetLysGluAlaLeu---------- 61

Qy  181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA 240
        -------------------------------------------------------------
Db   61 61

Qy  241 AAATACAATCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
        |||||||||||||||||||||||||||||||||||||||||||||
Db   62 -------LysValPheProThrValValSerThrSerPheIleGlnHisGluVal 77

Qy  301 GTGGAAGAGTATAGCCACCTGTTCACTATACCAAAGGCCTCGGACCCCAGCTTGCAGCCCTAC 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr 97

Qy  361 CTGCTGAATGGCTCCACTTTGATGTGGTCGCCTCCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 420
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db   98 LeuLeuMetAlaHisPheAspValAlaProGluGlyTrpGluValProPro 117

Qy  421 TTCTCTGGG 429
        |||||||||
Db  118 PheSerGly 120

RESULT 9
US-10-219-382A-2481
; Sequence 2481, Application US/10219382A
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Zhiwei
; APPLICANT: Wang, Gezhi
```

Right column

```
; APPLICANT: Ma, Yunqing
; TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
; FILE REFERENCE: 806A
; CURRENT APPLICATION NUMBER: US/10/219,382A
; CURRENT FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,261
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/488,725
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: US 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-01-25
; PRIOR APPLICATION NUMBER: US 09/491,404
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 09/496,914
; PRIOR FILING DATE: 2000-02-03
; PRIOR APPLICATION NUMBER: US 09/560,875
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/515,126
; PRIOR FILING DATE: 2000-05-18
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: pt_FL_genes Version 5.0
; SEQ ID NO 2481
; LENGTH: 120
; TYPE: PRT
; ORGANISM: Homo sapiens
DB: US-10-219-382A-2481

US-10-014-896-1 (1-1509) x US-10-219-382A-2481 (1-120)

Alignment Scores:
Pred. No.:
Score:                 1.9e-33        Length:       120
Percent Similarity:    562.50         Matches:      114
Best Local Similarity: 79.72%         Conservative: 0
Query Match:           79.72%         Mismatches:   0
                       20.37%         Indels:       29
                       6              Gaps:         1

Qy    1 ATGGCTCAGCGCGGTGCGTTTTGCGTTTGCTGCTGGCCTATGCTGCTCCTCCTCAGTTTTCCT 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    7 MetAlaGlnArgGlyValCysValLeuLeuLeuLeuValAlaMetLeuLeuLeuValPhePro 26

Qy   61 ACCGTCTCCAGATCGATGGGCCCGAGGAGGAGGCAATCAAAAGGCGTCGCGGAATCCCT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   27 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 46

Qy  121 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAAGAGGACGGCTGAAAGGTGCCATCCAG 180
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   47 SerGlnPheSerLysGluGluArgValAlaMetLysGluAlaLeu---------- 61

Qy  181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA 240
        -------------------------------------------------------------
Db   61 61

Qy  241 AAATACAATCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
        |||||||||||||||||||||||||||||||||||||||||||||
Db   62 -------LysValPheProThrValValSerThrSerPheIleGlnHisGluVal 77

Qy  301 GTGGAAGAGTATAGCCACCTGTTCACTATACCAAAGGCCTCGGACCCCAGCTTGCAGCCCTAC 360
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   78 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr 97

Qy  361 CTGCTGAATGGCTCCACTTTGATGTGGTCGCCTCCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 420
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db   98 LeuLeuMetAlaHisPheAspValAlaProGluGlyTrpGluValProPro 117

Qy  421 TTCTCTGGG 429
```

```
Db    118 PheSerGly 120
             ||||||||||
Qy        |||||||||| 
          AGAGGCGCTGAAAAGTGCCATCCAGAATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAA 215

RESULT 10
Sequence  9438, Application US/10156761
GENERAL INFORMATION:
  APPLICANT: OMURA, SATOSHI
  APPLICANT: IKEDA, HARUO
  APPLICANT: ISHIKAWA, JUN
  APPLICANT: HORIKAWA, HIROSHI
  APPLICANT: SHIBA, TADAYOSHI
  APPLICANT: SAKAKI, YOSHIYUKI
  APPLICANT: HATTORI, MASAHIRA
  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
  FILE REFERENCE: 249-262
  CURRENT APPLICATION NUMBER: US/10/156,761
  CURRENT FILING DATE: 2002-05-29
  PRIOR APPLICATION NUMBER: JP 2001-204089
  PRIOR FILING DATE: 2001-05-30
  PRIOR APPLICATION NUMBER: JP 2001-272697
  PRIOR FILING DATE: 2001-08-02
  NUMBER OF SEQ ID NOS: 15109
  SEQ ID NO 9438
  LENGTH: 440
  TYPE: PRT
  ORGANISM: Streptomyces avermitilis
US-10-156-761-9438

Alignment Scores:
Pred. No. :
Score:                   8.28e-17
Score:                   340.50
Percent Similarity:      44.49%
Percent Similarity:      28.19%
Best Local Similarity:   12.33%
Query Match:             6

Length:       440
Matches:      128
Conservative: 74
Mismatches:   198
Indels:       54
Gaps:         16

US-10-014-896-1 (1-1509) x US-10-156-761-9438 (1-440)
```

```
Qy   156 AGAGGCGCTGAAAAGTGCCATCCAGAATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAA 215
             |||||          |||         |||                 
Db    25 ArgGlyAlaGlyAspCysArg-GluArgProAlaAlaGluTyrValAlaGlnLys----- 42

Qy   216 TACTACAGCCCTGGCTGAGTTCGGAAAAATACATTCATAAAGTCTTTCCTACAGATGGTCAG 275
             ||||||||||:||||      |||    |||          |||::|    :::|:||||
Db    43 ------------LeuAlaGluLeuGly---                               54

Qy   276 CACCAGCTTTATCCAGCATCAAAGTCGTGAAGAGTATAGCCACCACGTTCACTATCCAAGG 335
             ||||::|    |:   |||           |||  |||||||||:|:          :::
Db    54 uArgThrGluGlyArgThrAsnValAlaAlaArg---                        68

Qy   336 CTCGGACCCAGCTTGCAGCCCTACCTGCTGGATAGGCTCACTTTGAATGTGGTGCCTGCCCC 395
             |||||          ||||||||    ||:|||||||||:|:|:||||||||||||||||
Db    68 yThrAspProLeuAlaAspAlaLeuLeuValHisLeuAspValValProAlaGl 88

Qy   396 TGAAGAAGGCTGGGAGGTGCCCCAATTCTCTAGGGTTGGAGGCGTGCATCTATGG 455
             :|||||||:||||||||||||::    |||:: |||||||::|||||    |||
Db    88 nAlaIleAspTrpSerValHisProPheSerGlyGluIleuArgAspGlyValTrpGl 108

Qy   456 TCGGGGCCACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAGGCCCTGGAGCTGCCT 515
             |||||          :::||||  ||||||||||||||||||||||||||::||||||||||
Db   108 yArgGlyAlaValAspMetLysAsn------                             123

Qy   516 GCTGATCAGGAAGTACATC----                                       560
             |||||||::||                                                 
Db   123 aValAlaArgAspTrpAlaArgProArgArgAspLeuValIleAlaPh 143

Qy   561 GGGCCATGATGAGGAGTCATCA------                                  605
             :::|                                                        
Db   143 eThrAlaAspGluGluAlaSerAlaValAspGlyPheGlnAlaAspArgHisPr 163

Qy   606 AGCCCTGCTCACAG------TCAAGGGGCGTCCAGCTAGCCTTCATTGTGGACGAGGGGGG 659
             |||                       |:||                            
```

```
Db   163 oGlyLeuPheGluGlyCysThrGluGlyIleSer-----------------GluSerGlyAlA 178
             ||||||||          |||        
Qy   660 CTTCATCTTGGATGATTGCATTCCTAACTCTGAGAAGAGCCCATCGCCCACTCTTCAGTCTC 719

Qy   178 aPheThrPheHisAsp---------GlyAlaGlyAspGlnLeuTyrProIleAlaAlaGl 195
             ||| 
Db   660 CTTCATCTTGGATGATTGCATTCCTAACTCTGAGAAGAGCCCATCGCCCACTCTTCAGTC

Qy   720 AGAGAAGGGTTCCATGAACCTCATGTCTTGCAAGTAAACATGACTTACAGGAGGACA 779
             ||||||:|:|||||        |||      :::      ||||||||:||||||||
Db   195 yGluArgGlyThrGlyTrpLeuTyrLeuAlaHisGlyArgAlaGlyHisGlySerLy 215

Qy   780 TCCTCCAAAGGAGACAACAGCCATTGGCATCCTTGCAGCTGCTGTCGTCAGGCAGTAGAC 839
             ||||||::|:||||||||||||||||||||:    |||      :::||||||||||||
Db   215 sValAsnArgSerAsnAlaValThrArgLeuAlaLeuAlaIleGlyAlaHi 235

Qy   840 ACCAATGCCTACATCATATTTGGAAGCGGGCACGTGGTGACTGTTATTGCAGCAACTGCAAA 899
             |||       ||||||||||||||||||||:    |||||||||:|:|||||||||||||
Db   235 sValAsnArgSerAsnAlaValThrArgLeuAlaLeuAlaIleGlyAlaHi 235

Qy   900 TGAGTTCCCTTCCCTGTCAATATA---ATCCTGAGCAACCCATGGCTATTGTGAACCACT 956
             ||| :::                 |||                   :::    :::||||
Db   254 aValTyrGlyIleAspProAlaAspLeuAspAspMetAspAspMetAspGlyIle 274

Qy   957 TATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACGACCACCGC 1016
             :::                   ||||||||||||||||||||:|::||||  :::|||||
Db   274 uLeuLysLysLeuGlyProAlaAlaLeuGluProThrValArgGlySerAlas 294

Qy  1017 ACTCACCCATATTCAAAAGCAGGGGTCATCAGTTCACGGCCTGGGTGACGTGGCCCAGGCCAC 1076
             |||||||          |||       |||||||:||||||||||          |||
Db   294 nProThrMetLeuserAlaGlyTyrTyrLysIleAsnValIleProGlyGluAlaValGlyTy 314

Qy  1077 AGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGACGAGA 1136
             ||| :::                     |||||||||:::              ||| 
Db   314 rValAspGlyArgTyrLeuProGlyGlyGluAspPheArgThrThrLeuAspArgProArgIle 334

Qy  1137 CATTGTGGCTGATAACAGAGTCCAGTTCATTGTGTTGAGTGTCTTTGACCCCTCCGT 1196
             ||||||          :::||||||||||||||          :::    |||
Db   334 uThrGlyProAspValGluTrpGluHisArgGlyLeuValAlaLeuLeuGlnAlaProVa 354

Qy  1197 CAGCCCTTCTGATGACAAGGCCTGGGCCTACCAGCCGCCGGACACCAGTCGT 1256
             |||||          :::    |||||||||||||||          ||| 
Db   354 lAsp---------------SerValThrYrAlaArgMetArgAlaValGluProPh 369

Qy  1257 CTTCCCGGAGTCAATATTACTGCCCCAGTTACTTCTATTGCGAACAGCACAGCGATT 1316
             ||||||          |||    |||||||||||||||          :::||| 
Db   369 eAlaProGluGlyHisVal---ValProTyrCysMetSerGlyThrGlyThrAspAlaLysGl 388

Qy  1317 CTTTACAAACCTCACCACTGGCATCCTACAGGTTCTACCCATCTACACTACAGCCTGAGA 1376
             |||||||||||||||||||||||||||:: |||||||||||          ||| 
Db   388 nPheSerArgLeuGlyIleThrGlyTyrGlyPheAlaProLeuLysSer---ProGluGl 407

Qy  1377 CTTTC----------AAACGCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTATGA 1427
             |||||               ||||||||:|||||||||||||||:|:||:||||||||||
Db   407 yPheAspTyrGlnAlaLeuProPheHisGlyValAspGlyUArgValProValGluAlaLeuHi 427

Qy  1428 GACCCAAGTGAAATTCATCTTTGAGTTGATTCAGAATGCT 1467
             |||||:::       |||      :::||||||||||||
Db   427 sPheGlyValArgValLeuAspArgPheLeuArgThrAla 440
```

Sequence 14162, Application US/10156761
GENERAL INFORMATION:
  APPLICANT: OMURA, SATOSHI
  APPLICANT: IKEDA, HARUO
  APPLICANT: ISHIKAWA, JUN
  APPLICANT: HORIKAWA, HIROSHI
  APPLICANT: SHIBA, TADAYOSHI
  APPLICANT: SAKAKI, YOSHIYUKI
  APPLICANT: HATTORI, MASAHIRA
  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
  FILE REFERENCE: 249-262
  CURRENT APPLICATION NUMBER: US/10/156,761
  CURRENT FILING DATE: 2002-05-29

```
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14162
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14162

Alignment Scores:
Pred. No.:
Score:                   3.94e-15
                         318.00
Percent Similarity:      43.08%      Length:         441
Best Local Similarity:   26.79%      Matches:        120
Query Match:             11.51%      Conservative:    73
DB:                      6           Mismatches:     159
                                     Indels:          96
                                     Gaps:            17

US-10-014-896-1 (1-1509) x US-10-156-761-14162 (1-441)

Qy   226 CTGGCTGAGTTCGGAAAATACATTCATTAAAGTCTTTCCTACAGTGGTTCAGTACCAGCTTT 285
             |||||||||||  |||||
Db    50 LeuAlaGluValGly------------------------------------------------  54

Qy   286 ATCCAGCATGAGACTCCGTCGGAAGAGTAT------------------------AGCCACCTGTTCACTATCCAA 333
Db    55 LeuGluProGlnIlePheGluSerHisGlnIleGlyArgAlaSerThrValAlaAlaArgIleGlu  74

Qy   334 GGGTCCGGACCCAGCTTGCACCCTACCTGGCTCACTTGATGTGTGGCCTGCC 393
Db    75 GlyGluAspProSerArgProAlaLeuLeuIleHisGlyHisThrAspValValProAla  94

Qy   394 CCTGAAGAGAAGGCTGGGAGGGTGCCCCAATTCTCTGGGTTGGGAGGCGATGGGGTCATCTAT 453
Db    95 AsnAlaLaAspTrpTrpPheThrHisHisProPheSerGlyValIleAlaAlaSpGlyCysValTrp 114

Qy   454 GGTGGGGGCCACTGGACCAAGAACTCTGTGCATGGCATTACTGCCAGGCCTTGGAGCTC 513
Db   115 GlyArgGlyIleAlaValAspMetLysAsp------------------------ 129
                                         MetAspAlaMetThrLeu

Qy   514 CTGCTGATCAGGAGAAGTACATC-------------------CCCCGAAGATCTTTCTTCATTTCT 558
Db   130 AlaLaValArgAspArgLeuArgThrGlyArgArgProProArgAspIleValLeuAla 149

Qy   559 CTGGGCCATGATGAGGGACTCATCAGGGACA--GGGGCTCAGAGGANTCTCAGCCCTGCTA 615
Db   150 PheLeuAlaAspGluGluAlaGlyGlyThrYrGlyValAlaLys--------TyrLeuVal 166

Qy   616 CACTCA------AGGGGCGGTCCAGCTAGCCCTTCATTGTGGACGAGGGGGGC 660
Db   167 GlnLysHisProAspLeuProHisGluGlyIleValThrGluAla---IleGlyLeuValGlyGly 185

Qy   661 TTCATCTTGGATCATTTCATTCCTAACTTCAAGAAGCCATCGGCCCTGATTGCAGTCTCA 720
Db   186 Phe--------SerPheThrValAsnGlyLysArgLeuTyrLeuValGluThrArgAla 202

Qy   721 GAGAGAGGGTTCCATGAACCTCATGCCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCT 780
Db   203 GlnLysGlyMetHisTrpMetArgLeuThrValAspGlyHisIleGlySerMet 222

Qy   781 CCTCCAAAGGAGACAAGCATTGGCATCCTTCGCACGTCCTGTGCAGCCGGATTGGGAGCAGACA 840
Db   223 ThrAsnAspAspAsnAlaIleThrGluCysGluAlaValGlyLeuGlyArgHis 242

Qy   841 CCAATGCCTATCATATTTGGAACGCGGACAGTGGTGACTGTATTTGCAGCAACTGGCAAAT 900
Db   243 ThrTrpProValArgVal---ThrLysThrValArgSerPheLeuAspGluLeuSerAsp 261

Qy   901 GAGTTTCCCTTCCCTGTCAAATAATAATCCTGAGCAACCCATGCTATTTGAACCACTATA 960
Db   262 AlaLeuGlyThrGluLeu----------- 276

Qy   961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGAGACCACCACGGCACTC 1020
Db   277 AlaLysLeuGlyIleAlaLysMetValGlyAlaThrLeuArgAsnSerAlaAlaPro 296

Qy  1021 ACCATTATTCAAAGCAGAGGGGTCAAGTTCAAATGTCATCCCCCCAGTGGCCCAGGCCACGTC 1080
Db   297 ThrMetLeuGlyAlaGlyTyrLysValAsnValIleProGlyGlnAlaThrAlaHisVal 316

Qy  1081 AACTTCCGGATTCACCCTGACAGACAGAGTCCAAGAGGGTCCTAGAACTCACGAAG---- 1134
Db   317 AspGlyArgPheLeuProGlyHisGluGluGluProLeuLeuAlaAspLeuAspArgIleLeu 336

Qy  1135 ----------------------------------AACATTGTGGCTGATAACAGAGTCCAGTTC--- 1164
Db   337 GlyProArgValLysGluAspValHisGlyAspLysAlaLeuGluThrAspPheAsp 356

Qy  1165 -----------------------CAGTGTTGAGTGCCTTTGACCCCTCCCCCGTCAGC 1200
Db   357 GlyArgLeuValAlaSpAlaMetGlnSerAlaLeuSerAlaLeuGluAspProIleAla 374

Qy  1201 CCTTCTGATGACAAGGGCCTTTGGGCTACCAGCTGCTCCGCGCAGACCGTACAGTCCGTCTTC 1260
Db   375 --------LysAlaValProTyrMetLeu-------------- 381

Qy  1261 CCGGAAGTCAAATATTACTGCCCCAGTTACTTCTTATTGGCAACACAGACAGCAGCCGGATTCTTT 1320
Db   382 -----------SerGlyIleGlyThrAspAlaLysSerPhe 390

Qy  1321 ACAAACCTCACCACTGGCATCTACAGGTTCTCACCCATCTACACACACGACCTGAA---GAC 1377
Db   391 AspAspLeuGlyIleArgCysPheGlyPheAlaProLeuLysLeuProGluLeuAsp 410

Qy  1378 TTCAAA---CGCATCCATGGAGTCAACGAGAAAAATCTCAGTTGCAGCCTATGAGACCCAA 1434
Db   411 PheLeuaGlyMetPheHisGlyIleAlaSpLeuArgValProGlyLeuLysPheGly 430

Qy  1435 GTGAAATTCATCTTTGAGTTGATT 1458
Db   431 ValArgValLeuAspArgPheIle 438
```

```
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: pt_FL_genes Version 5.0
;
; SEQ ID NO 1141
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-219-382-1141

Alignment Scores:
Pred. No.:
Score:                     1.17e-13
Percent Similarity:        66.94%    Matches:      75
Best Local Similarity:     61.98%    Conservative:  6
Query Match:               10.75%    Mismatches:   22
DB:                            6     Indels:       18
                                     Gaps:          5

US-10-014-896-1 (1-1509) x US-10-219-382-1141 (1-112)

Qy      1 ATGGGTCAGGCGGTGCGGTTGCGTGCTGCCTGGCCCTGGGGGCTATGCTCCTCCTAGTTTTCCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGGTCAGGCGGTGCGGTTGCGTGCTGCCTGGCCCTGGGGGCTATGCTCCTCCTAGTTTTCCCT 60

Qy      1 MetAlaGlnArgCysValCysValLeuLeuAlaLeuMetLeuLeuLeuLeuValPhePro 20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MetAlaGlnArgCysValCysValLeuLeuAlaLeuMetLeuLeuLeuLeuValPhePro 20
```

(table continued — alignment listing)

```
; PRIOR FILING DATE: 2000-02-28
; PRIOR APPLICATION NUMBER: US 09/577,409
; PRIOR FILING DATE: 2000-05-18
; PRIOR APPLICATION NUMBER: US 09/519,705
; PRIOR FILING DATE: 2000-03-07
; PRIOR APPLICATION NUMBER: US 09/574,454
; PRIOR FILING DATE: 2000-05-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2818
; SOFTWARE: pt_FL_genes Version 5.0
;
; SEQ ID NO 1141
; LENGTH: 112
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-219-382A-1141

Alignment Scores:
Pred. No.:
Score:                     1.17e-13
Percent Similarity:        66.94%    Matches:      75
Best Local Similarity:     61.98%    Conservative:  6
Query Match:               10.75%    Mismatches:   22
DB:                            6     Indels:       18
                                     Gaps:          5

US-10-014-896-1 (1-1509) x US-10-219-382A-1141 (1-112)
```

; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932126.4
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932127.2
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932128.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932129.9
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: DE 19932226.0
; PRIOR FILING DATE: 1999-07-09
; PRIOR APPLICATION NUMBER: DE 19932920.6
; PRIOR FILING DATE: 1999-07-14
; PRIOR APPLICATION NUMBER: DE 19932922.2
; PRIOR FILING DATE: 1999-07-14
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 442
; SEQ ID NO 202
; LENGTH: 422
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-10-454-437-202

Alignment Scores:

Pred. No.:	
Score:	7.47e-12
Percent Similarity:	274.00
Best Local Similarity:	42.55%
Query Match:	25.48%
DB:	9.92%
	6

Length:	422
Matches:	106
Conservative:	71
Mismatches:	161
Indels:	78
Gaps:	20

US-10-014-896-1 (1-1509) x US-10-454-437-202 (1-422)

```
Qy   313  AGCCACCTGTTCACTATCCAAGGCTCGGACCCCCAGCTTGCACCCTACCTGCTGATGGCT   372
               ::|||  ||| |||::|||||||||   |  |||||| ||:||  |||||
Db    41  SerIleIleValThrValProGlySerAspSerProAspProAlaGlnProLeuThrLeuGly  60

Qy   373  CACTTTGATGTGGTGCCCTGCCCCCTGAAGAAGGCTGGCCCCCCATTCTCTGGGGTTG   432
               ||| :: |||  :::|| |||||||| |||  | |||::||| ||| :||:||
Db    61  HisThrAspValProValAspLeuProLysTrpThrLysAspProHisGlyAlaGluGlu  80

Qy   433  GAGGGTGATGGCGTCATCTATGGTCGGGGACACACTGGACACAAGAACTCTGGATGGCCA   492
          ||||||  |||||||||||||::|||||||||||||||||||||:||||||  ||||||
Db    81  IleSerAspGlyValIleTyrGlyArgGlyHisThrGlyThrArgAsnSerGlyTrpAla  95

Qy   493  TTACTGCAGGCCTTGGAGCTCCTGCTCACGAAGTACAATCCCCCGAGCGATCT-----   546
          :::  |||  |||   ||||::|||||| ||||||   |||||| |||||:|
Db    96  PheIleThrAlaThrGlnAlaLeuAlaValThrArgGln---ValAlaArgGlyGluGlyLeu  114

Qy   547  --------TTCTTCATTTCTCTGGGCCATGATGAGGAGTCATCAGGGACAGGGGCT   594
                  :|:|   ||| :||:: |||  ||||||   |||::||
Db   115  ArgGlyThrLeuThrPheValGlyValAlaAspGluGluAlaAspGlyValAlaAla  134

Qy   595  CAGAGGATCTCA----------------------GCCCTGCTACAGTCA   621
          :::           :::|||
Db   135  LysTrpLeuSerGluGluHisGlnAsnLeuPheSerTrpLysAsnCysLeuSerGluSer  154

Qy   622  AGGGGCGGTCCAGCTAGCCTTCATTGTGGGACGGGGGGTTCATCTTGGATGATTTCATT   681
          ||| ||| |||  ::|||
Db   155  GlyGlySerHisLeuPro-----ValHisAspGly  164

Qy   682  CCTAACTTCAAGGAAGCCCATCCGCTTGATTGGCAGCTCAGAGAAGGGTTCCATGAACCTC   741
          ||| ||| |||  ::|||
Db   165  -------SerAspAlaValLeuAsnValGlyGluLysGlyAlaAlaGlnArg  180

Qy   742  ATGCTGCAAGTAAACATGACTTCAGGCCCACTCTTCAGCTCCT---CCAAAGGAGAGCAAGC   798
          :::  |||  |||  :||||   ||||||  |||
Db   181  ArgIleHisValAsnGlyAspAlaGlyHisGlySerIleProPheAspAspSerAla  200

Qy   799  ATTGGCATCCTTGCAGCTGCTGTGCAGCCGCCGAATTGGAGCAGAGACACCAATGCCTATCATA-   855
          |||       ::|
Db   201  IleValLysIleGlyLysGluValAlaAlaArgArgIleAlaAlaAlaAspLeuLysValAlaLys  220
```

; SEQ ID NO 2506
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Propioni acnes
PCT-US02-32727-2506

RESULT 15
PCT-US02-32727-2506
; Sequence 2506, Application PC/TUS0232727
; GENERAL INFORMATION:
; APPLICANT: Mitcham, Jennifer
; APPLICANT: Skelky, Yasir
; APPLICANT: Persing, David
; APPLICANT: Bhatia, Ajay
; APPLICANT: Maisonneuve, Jean Francois
; APPLICANT: Zhang, Yanni
; APPLICANT: Wang, Siqing
; APPLICANT: Jen, Shyian
; APPLICANT: Lodes, Michael
; APPLICANT: Benson, Darin
; APPLICANT: Jones, Robert
; APPLICANT: Carter, Darrick
; APPLICANT: Barth, Brenda
; APPLICANT: Douglass, John
; TITLE OF INVENTION: Compositions and Methods for the Therapy and Diagnosis of Acne
; FILE REFERENCE: 210121.514C1
; CURRENT APPLICATION NUMBER: PCT/US02/32727
; CURRENT FILING DATE: 2002-10-11
; NUMBER OF SEQ ID NOS: 30092
; SEQ ID NO 2506

```
Qy   856  --------TTTGAAGCGGGGACAGTGGGTGACTGTTGACCAGCAACTGGGCAAATGAGTTT---   906
                  |||:::                    |||:::          |||
Db   221  AspAspIleTrpGlnGlyPheVal------------------GlnAlaHisArgPheAsp  234

Qy   907  CCCTTCCCTGTCAAATATAATAATCCTGAGC-----AACCCATGGCTATTTGAACCACTT---   957
          |||:::|||         :::|||||        |||:::|||
Db   235  ProGluThrGluGlnAlaLeuLeuSerGlyThrSerProGluAlaTyrAlaGluPheGly  254

Qy   958  ---ATAAGCAGGTTTATGGAGAGAGAAAATCCCTTAACCAATAATCAGGACCACCACG   1014
             :::|||||||         |||:::       |||
Db   255  GlyLeuSerArgPhe--------------AlaHisAlaValSerHisLeuThrIle  268

Qy  1015  GCACTCACCATATTCAAAGCAGAGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCC   1074
          |||  |||||||   |||:::|||         |||:::
Db   269  AlaGlnThrValValArgAlaGlyIleAlaIleAsnValLeuProSerHisAlaTyrLeu  288

Qy  1075  ACAGTCAACTTCCGGATTCACCCCTGACAGACA----------GTCCAAGAGGTCCTAGAA   1125
          |||:::|||||||||||||||          |||:::
Db   289  GlyLeuAspIleArgThrLeuProGlyGlnThrAsnAspThrValAspAspThrLeuArg  308

Qy  1126  CTCACGAAGAAACAATTGTGGCTGATAACAGATCCACAGTTCCATGTGTTGACTGCCTTTGAC   1185
          |||:::||||||          :::
Db   309  AlaAlaAlaLeuGlyAspLeuAspGluValIleGluIleLeuIleSer-----Glu  326

Qy  1186  CCCCTCCCGTCAGCCCTTCTTGATGACAAGGCCGTGGGCGTCTACAGCTCTCCCCAGACC   1245
          |||||||||||||||||||:::|||   ||:::|||
Db   327  GluAlaThrValSerProThrAspSerAspArgLeu--------TyrAsnThrLeuGluLysVal  344

Qy  1246  GTACAGTCCGTCTTCCCGGAGTCAATATTACTGCCCGCCAGTTACTTCTTATTGGCAACACA   1305
          |||:::|||||||||||||||          :::|||
Db   345  LeuGlyAspPhePheProAspAlaProVal---ValProIleIleSerSerGlyGlySer  363

Qy  1306  GACAGCCCATTCTTTACAAAACCTCACCACTGGCACTCTACAGGTTCTCACCCCCACTACATA   1365
          |||:::|||                    :::|||
Db   364  AspLeuArgPheGlyArgArgLeu----GlyGlyValGlyTyrGlyPheAlaValHisAla  382

Qy  1366  CAG-------------CCTGAAGACTTCAAACGACATCCATGGAGTCAACGAG------   1404
          |||                  |||:::|||
Db   383  ArgGluArgThrLeuAlaGluLeuAlaMetGlyLeuLeuHisSerHisAspGluAlaLeuTyr  402

Qy  1405  ----------AAAATCTCAGTCCAAGCCTTATGAGACCCCAAGTGAAA   1440
                    :::|||
Db   403  LeuGlyAspLeuGluLeuThrValArgGlyTyrAspSerValValArg  418
```

```
Alignment Scores:
Pred. No.:
Score:                2.74e-11        Length:             458
Percent Similarity:   266.50          Matches:            117
Best Local Similarity: 41.32%         Conservative:        71
Query Match:          25.71%          Mismatches:         212
DB:                   9.65%           Indels:              55
                      1               Gaps:                18

US-10-014-896-1 (1-1509) x PCT-US02-32727-2506 (1-458)

Qy   175  ATCCAGATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAAT------------------  216
          ::::||||||||            ::::
Db    35  ILeGlnILeAspSerGlnAsnPheGlyProGlnAsnAlaArgGlyGlyGluMetCys   54

Qy   217  ------ACTACAGCCCCTGCCTGAG-----------------------------------  261
              ||    :||              ::::::||||
Db    55  HisTyrValThrGlyLeuLeuAspGluILeGlyValGlyValThrLeuHisGluSerGlu  74

Qy   262  CCTACAGTGGTCAGCACCACAGCTTTATCCAGCATCGAAGTCGTGGAAGAGTATAGCCACCTG  321
          |||   ||||||::            ||||
Db    75  ProGlyArgValThr-------             LeuValAlaGluTrpAlaPro---  86

Qy   322  TTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGAT  381
          ::||||::||||            |||||
Db    87  -----GluGlyThrAspThrSerArgProAlaLeuLeuLeuHisGlyHisSerAsp  103

Qy   382  GTGGTGCCTGCCCTGAGAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTGGAGCGTGAT  441
          |||||||||            |||
Db   104  ThrValProPheGluAlaAlaAspTrpThrThrHisHisProLeuSerGlyILeHisAsp  123

Qy   442  GGGCGTCATCTATTGGTTCGGGGGCCACAGCGGACGAACAAGAACTCTGTGATGGCCATTACTGCGAG  501
          :::||||            :||||| ||:
Db   124  AsnCysValTrpGlyAlaArgGlyAlaILeAspMetLysGlyPheLeuAlaMetValLeuSer  143

Qy   502  GCCTTGGAGCTCCTGCCTGATCAGAGAAGTACATCCCCGAAGATCTTTCTTCATTTCTCTG  561
          ::|||:::||||            |||:||| |||
Db   144  AlaILeArgAlaArgGlnAsnArgGlyLysGlnValProSerArgProILeArgPheILeMet  163

Qy   562  GGCCAATGATGAGGAGGAGTCATCAGGGACA----------------GGGGCTCAGGGATC  603
          :||||            ||:
Db   164  PheAlaAspGluGluGluSerGlyTyrThrLeuGlySerThrTrpLeuGlyAlaThrHisPro  183

Qy   604  TCAGCGCCTGCTACAGTCAAAGGGGGGGGCTCCAGCTAGCCTTCATTGTGGACCAGAGGGGGGGGCTTC  663
          ::||||            ::|||||||||
Db   184  GluAlaPhe---------AspGlyValThrGlyAla---ILeSerGluValGlyGlyPhe  199

Qy   664  ATCTTGGATGATTTCATTCCTAACTTCACAGAAGCCCATCGCCTTGATCGAGTCTCAGAG  723
          |||            :||| |||
Db   200  SerLeuThr--------ThrProGlnGlyLysArg---ValTyrValILeGlnSerAlaGlu  216

Qy   724  AAGGGTTCCATTGAACCTCATGCTCGAAGTAAACATGACTTCAGCGCCACTCTTCAGCTCCT  783
          |||||            :::                    :::
Db   217  LysGlyLeuTrpILeTrpThrPheArgMetSerALaThrGlySerThrHisGlySerMetArg  236

Qy   784  CCAAAGGAGAACAAGCAACATTGGCATCCTTGGCTGCTGCTGGACCGAGACGAGACACCA  843
          |||            :::|||||||||:::
Db   237  AsnProAspAsnAlaValThrArgValLeuAspALaLeuSerArgILeAspSerTyrGln  256

Qy   844  ATGGCTTATCATATTTGGACGGGGACAGTGGTGCTGACTGTATTGCACCAAACTGGCAAATGAG  903
          |||            :::
Db   257  TrpProAspLeuHisHis------------ProValGlnGluPheLeuAsnGln  271

Qy   904  TTTCCCTTCCCTGTCAAATATAAATCCTGAGCAACCCATGGCTATTTGAACCACACTTTATAAGC  963
          |||            :::
Db   272  ValAlaALaMetTrpGlyLeuThrILeAspArgAspAspLeu----GluSerSerLeuSer  290

Qy   964  AGGTTTATGGAGAGAGAAATCCCTTACCAATGCAATAAATCAGGACCACACCAGGGCACTACC  1023
          |||            :::
Db   291  ProILeGlySerLeuSerArgMetValALaALaCysTyrALaHisAsnValThrProThr  310

Qy  1024  ATATTCAAAGCAGGGGTCAAGTTCAAATGTCATCCCCCAGTGGGCCCCAGCGCCACAGTCAAC  1083
          :::||||||            |||||||::::|||
Db   311  ValLeuSerALaGlyTyrLysValAsnValProThrArgALaSerArgALaGlyValAlaAsp  330
```

```
Qy  1084  TTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACAATGTG  1143
          |||            ::::||||||            :::||||
Db   331  AlaArgPheILeProGly-----------------AlaGluGluAspMetILeLeuThrILeLysSerLeu  348

Qy  1144  GCTGATAACAGAGATCCAGTTCCAATGTGTTGAGTGCCTTGACCCCCTCCCCGTCAGCCCT  1203
          |||            :::||||            |||
Db   349  AlaGlyProGlyILeAspPheGluThrILeSerArg---LysProAlaThrALaALaPro  367

Qy  1204  TCTGATGACAAGGCCTTGGGCTACCAGCTGTCCCGCCAGACCGTACAGTGCCGTCTTCCCG  1263
          :::||||||            :::||||            |||
Db   368  PheGluGlyALaALaVal-----AspALaILeAlaArgGlnALaValAspSerALaGluAspPro  385

Qy  1264  GAAGTCAATATTACTGCCCCAGTTACTTCTCTATTGGCAACACAGACCAGCCGGATTCTTTACA  1323
          ||||||            :::
Db   386  GlyALa---ValValLeuProTyrLeuAsnSerALaGlyThrAspALaALaGlyPheILeA  404

Qy  1324  AACCTCACCACTGGC------------ATCTACAGGTTCTACCCCCATCTACTACATACAGCCT  1371
          |||            ||||||
Db   405  ValLeuProAspGlyArgArgILeAsnTyrTyrGlyCysThrProLeueuArgLeu---Pro  423

Qy  1372  GAAGACTTCAAACGGCATC----------CATGGAGTCAACGAGAAAATCTGAGTCCAAGCC  1422
          |||||| ||            :::
Db   424  AlaAspPheAspPheILeAsnLeuPheHisGlyValAlaAspGluArgValProValGlySer  443

Qy  1423  TATGAGACCCAAGTGAAAATTCATCTTTGAGTTGATTCAGAAATGCT  1467
          |||||            :::
Db   444  LeuValPheGlyALaLysValValAspHisILeLeuGlnGluAla  458

Search completed: June 27, 2003, 11:23:51
Job time : 175.5 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:52:17 ; Search time 3668 Seconds
                                            (without alignments)
                                            10343.515 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1 atggctcagcgggtgcggtttg...........ctcacctgcacaaactgtga 1509

Scoring table: IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched: 24791104 seqs, 12571243825 residues

Total number of hits satisfying chosen parameters: 49582208

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Database :   Pending_Patents_NA_Main:*

 1: /cgn2_6/ptodata/1/pna/PCTUS_COMB.seq:*
 2: /cgn2_6/ptodata/1/pna/US06_COMB.seq:*
 3: /cgn2_6/ptodata/1/pna/US07_COMB.seq:*
 4: /cgn2_6/ptodata/1/pna/US080_COMB.seq:*
 5: /cgn2_6/ptodata/1/pna/US081_COMB.seq:*
 6: /cgn2_6/ptodata/1/pna/US082_COMB.seq:*
 7: /cgn2_6/ptodata/1/pna/US083_COMB.seq:*
 8: /cgn2_6/ptodata/1/pna/US084_COMB.seq:*
 9: /cgn2_6/ptodata/1/pna/US085_COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US086_COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US087_COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US088_COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US089_COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US090_COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US091_COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US092_COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US093_COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US094_COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US095A_COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US095B_COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US095C_COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US095D_COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US096A_COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US096B_COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US096C_COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US096D_COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US096E_COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US097A_COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US097B_COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US097C_COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US098A_COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US098B_COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US098C_COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US099A_COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US099B_COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US099C_COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US099D_COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US100A_COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US100B_COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US101A_COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US101B_COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US102A_COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US102B_COMB.seq:*
44: /cgn2_6/ptodata/1/pna/US6000_COMB.seq:*
45: /cgn2_6/ptodata/1/pna/US6001_COMB.seq:*
46: /cgn2_6/ptodata/1/pna/US6002_COMB.seq:*
47: /cgn2_6/ptodata/1/pna/US6003_COMB.seq:*
48: /cgn2_6/ptodata/1/pna/US6004_COMB.seq:*
49: /cgn2_6/ptodata/1/pna/US6005_COMB.seq:*
50: /cgn2_6/ptodata/1/pna/US6006_COMB.seq:*
51: /cgn2_6/ptodata/1/pna/US6007_COMB.seq:*
52: /cgn2_6/ptodata/1/pna/US6008_COMB.seq:*
53: /cgn2_6/ptodata/1/pna/US6009_COMB.seq:*
54: /cgn2_6/ptodata/1/pna/US6010_COMB.seq:*
55: /cgn2_6/ptodata/1/pna/US6011_COMB.seq:*
56: /cgn2_6/ptodata/1/pna/US6012_COMB.seq:*
57: /cgn2_6/ptodata/1/pna/US6013_COMB.seq:*
58: /cgn2_6/ptodata/1/pna/US6014_COMB.seq:*
59: /cgn2_6/ptodata/1/pna/US6015_COMB.seq:*
60: /cgn2_6/ptodata/1/pna/US6016_COMB.seq:*
61: /cgn2_6/ptodata/1/pna/US6017_COMB.seq:*
62: /cgn2_6/ptodata/1/pna/US6018_COMB.seq:*
63: /cgn2_6/ptodata/1/pna/US6019_COMB.seq:*
64: /cgn2_6/ptodata/1/pna/US6020_COMB.seq:*
65: /cgn2_6/ptodata/1/pna/US6021_COMB.seq:*
66: /cgn2_6/ptodata/1/pna/US6022_COMB.seq:*
67: /cgn2_6/ptodata/1/pna/US6023_COMB.seq:*
68: /cgn2_6/ptodata/1/pna/US6024_COMB.seq:*
69: /cgn2_6/ptodata/1/pna/US6025_COMB.seq:*
70: /cgn2_6/ptodata/1/pna/US6026_COMB.seq:*
71: /cgn2_6/ptodata/1/pna/US6027_COMB.seq:*
72: /cgn2_6/ptodata/1/pna/US6028_COMB.seq:*
73: /cgn2_6/ptodata/1/pna/US6029_COMB.seq:*
74: /cgn2_6/ptodata/1/pna/US6030_COMB.seq:*
75: /cgn2_6/ptodata/1/pna/US6031_COMB.seq:*
76: /cgn2_6/ptodata/1/pna/US6032_COMB.seq:*
77: /cgn2_6/ptodata/1/pna/US6033_COMB.seq:*
78: /cgn2_6/ptodata/1/pna/US6034_COMB.seq:*
79: /cgn2_6/ptodata/1/pna/US6035_COMB.seq:*
80: /cgn2_6/ptodata/1/pna/US6036_COMB.seq:*
81: /cgn2_6/ptodata/1/pna/US6037_COMB.seq:*
82: /cgn2_6/ptodata/1/pna/US6038_COMB.seq:*
83: /cgn2_6/ptodata/1/pna/US6039_COMB.seq:*
84: /cgn2_6/ptodata/1/pna/US6040_COMB.seq:*
85: /cgn2_6/ptodata/1/pna/US6041_COMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1509	100.0	1509	38	US-10-014-896-1	Sequence 1, Appl
2	1505.8	99.8	1509	78	US-60-345-785-3	Sequence 3, Appl
3	1505.8	99.8	2229	78	US-60-345-785-1	Sequence 1, Appl
4	1014	67.2	1086	38	US-10-014-896-3	Sequence 3, Appl
5	860.4	57.0	932	25	US-09-648-365-17	Sequence 17, Appl
6	860.4	57.0	932	42	US-10-259-776-17	Sequence 17, Appl
7	673.4	44.6	1271	26	US-09-666-026-75	Sequence 75, Appl
8	673.4	44.6	1271	59	US-60-156-625-265	Sequence 265, App
9	596.8	39.5	909	80	US-60-360-207-413	Sequence 413, App
10	582	38.6	1307	1	PCT-US02-09921-430	Sequence 430, App
11	572.8	38.0	722	25	US-09-648-365-18	Sequence 18, Appl
12	572.8	38.0	722	42	US-10-259-776-18	Sequence 18, Appl
13	473.4	31.4	649	65	US-60-213-362-6769	Sequence 6769, Ap
14	473.4	31.4	649	69	US-60-250-830-1230	Sequence 1230, Ap
15	473.4	31.4	649	76	US-60-323-956-1230	Sequence 1230, Ap
16	403.6	26.7	.706	65	US-60-213-845-615	Sequence 615, App
17	359.4	23.8	2432	80	US-60-360-207-412	Sequence 412, App
18	312	20.7	312	21	US-09-540-733-6210	Sequence 6210, Ap
19	297	19.7	308	21	US-09-540-733-5997	Sequence 5997, Ap
20	285.2	18.9	391	69	US-60-252-833-40714	Sequence 40714, A
21	267	17.7	319	13	US-08-904-468-45	Sequence 45, Appl

22	267	17.7	319	21	US-09-547-599C-1191	Sequence 1191, Ap
23	245	16.2	362	1	PCT-US02-25485-2122	Sequence 2122, Ap
24	245	16.2	362	16	US-09-289-768-12086	Sequence 12086, A
25	245	16.2	362	19	US-09-515-126-5026	Sequence 5026, Ap
26	245	16.2	362	22	US-09-577-409-5026	Sequence 5026, Ap
27	245	16.2	362	35	US-09-939-397-12086	Sequence 12086, A
28	245	16.2	362	75	US-60-311-261-2114	Sequence 2114, Ap
29	239.4	15.9	333	21	US-09-540-213-22115	Sequence 22135, A
30	239.4	15.9	333	58	US-60-142-699-126	Sequence 126, App
31	239.4	15.7	680	63	US-60-195-044-82	Sequence 82, Appl
32	236.8	15.7	3561	65	US-09-213-845-85	Sequence 85, Appl
33	236.4	15.7	943	24	US-09-634-308B-147337	Sequence 147337,
34	236.4	15.7	943	38	US-10-027-632-147337	Sequence 147337,
35	235.8	15.6	255	21	US-09-540-733-7044	Sequence 7044, Ap
36	216	14.3	288	21	US-09-540-210B-34061	Sequence 34061, A
37	199.2	13.2	283	20	US-09-535-896-5429	Sequence 5429, Ap
38	197.2	13.1	281	62	US-60-184-777-1196	Sequence 1196, Ap
39	172.2	11.4	339	1	PCT-US02-25485-99	Sequence 99, Appl
40	172.2	11.4	339	75	US-60-311-261-99	Sequence 99, Appl
41	172	11.4	294	14	US-09-062-736-3121	Sequence 3121, Ap
42	172	11.4	294	21	US-09-540-212A-11876	Sequence 11876, A
43	172	11.4	294	48	US-60-045-650-3121	Sequence 3121, Ap
44	167	11.1	258	15	US-09-113-753-949	Sequence 949, App
45	167	11.1	258	21	US-09-540-212A-66311	Sequence 66311, A

ALIGNMENTS

```
RESULT 1
US-10-896-1
Sequence 1, Application US/10014896
; Sequence 1, Application US/10014896
; GENERAL INFORMATION:
; APPLICANT: Yu, Xuanchuan
; APPLICANT: Miranda, Maricar
; APPLICANT: Friddle, Carl Johan
; TITLE OF INVENTION: Novel Human Proteases and
; TITLE OF INVENTION: Polynucleotides Encoding the Same
; FILE REFERENCE: LEX-0280-USA
; CURRENT APPLICATION NUMBER: US/10/014,896
; CURRENT FILING DATE: 2001-12-11
; PRIOR APPLICATION NUMBER: US 60/255,567
; PRIOR FILING DATE: 2000-12-14
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 1509
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-014-896-1

Query Match      100.0%;    Score 1509;    DB 38;    Length 1509;
Best Local Similarity    100.0%;
Matches 1509;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;
```

```
Qy      1 ARGGCTCAGCGGGTGCGTTGCGTGCGTGGCCTGGCCCTTGGTGTGGCTTATGCTGGCCTAGTTTCCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ARGGCTCAGCGGGTGCGTTGCGTGCGTGGCCTGGCCCTTGGTGTGGCTTATGCTGGCCTAGTTTCCCT 60

Qy     61 ACGGTCTCCAGATCGATGGGCCCGAGGACGGGGAGCATCAAAGGGCGTCCGGGAAATCCCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACGGTCTCCAGATCGATGGGCCCGAGGACGGGGAGCATCAAAGGGCGTCCGGGAAATCCCT 120

Qy    121 TCTCAGTTCACCAAAGAGAAACGGCCGGATGAAAGAAGGGTGCCATCCAG 180
          ||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTCAGTTCACCAAAGAGGAACGGCTCGGCATGAAAGAAGGTGCCCATCCAG 180

Qy    181 ATTCCAACAGTGACTTTTAGCTCTCGAGAAGTCCAATACTACACGCCCTGGCCGTCAGTTCGGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATTCCAACAGTGACTTTTAGCTCTCGAGAAGTCCAATACTACACGCCCTGGCCGTCAGTTCGGA 240

Qy    241 AAATACATTCATTAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATACATTCATTAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300

Qy    301 GTGGGAAGAGTATAGCCACCTTGTTCACTATCCCAAGGCGTCGGACCCCAGCCAGCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGGAAGAGTATAGCCACCTTGTTCACTATCCCAAGGCGTCGGACCCCAGCCAGCCTAC 360

Qy    361 CTGGCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGGCTGGAAGGTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGGCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGGCTGGAAGGTGCCCCA 420

Qy    421 TTCTCTGCGGGGTTGGGAGCCGGTGAATGGGCGTCATCTATGGGTCGGGGCACACTCGGACGACAAGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTCTCTGCGGGGTTGGGAGCCGGTGAATGGGCGTCATCTATGGGTCGGGGCACACTCGGACGACAAGAAC 480

Qy    481 TCTGTGATGGGCATTACTGCAGGGCTCCTGCTGATCAGGAAGGTACATCCCCGA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTGTGATGGGCATTACTGCAGGGCTCCTGCTGATCAGGAAGGTACATCCCCGA 540

Qy    541 AGATCTTTCTTCATTTCTCTGGGCCGCAGTCAGTGGTGGGCTCAGAGG 600
          |||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGATCTTTCTTCATTTCTCTGGGCCCATGAGTGATGAGGACCAGGGGCTCAGAGG 600

Qy    601 ATCTCAGCCCCTGCTACAGTCAAAGGGGGCGTCCAGCTAGCCTTCATTGTGTGGAGGGGGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCTCAGCCCCTGCTACAGTCAAAGGGGGCGTCCAGCTAGCCTTCATTGTGTGGAGGGGGC 660

Qy    661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGATTAAACATTCAGCCCACTCTTCAGCT 720
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGATTAAACATTCAGCCCACTCTTCAGCT 720

Qy    721 GAGAAGGGTTCCATGAAACCTCATCCTGCGAAGTAAACATGACTTCAGCGCCACTCTTCAGCT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAGAAGGGTTCCATGAAACCTCATCCTGCGAAGTAAACATGACTTCAGCGCCACTCTTCAGCT 780

Qy    781 CCTCCAAAAGGACAACAAGCAATTGGCATCCCTTGCACGCTGCTGTCAGCCGATTGGAGCAGACA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCTCCAAAAGGACAACAAGCAATTGGCATCCCTTGCACGCTGCTGTCAGCCGATTGGAGCAGACA 840

Qy    841 CCAATGCCTATCATATTTGGGAAGAGCAGTGGTGACTGTATTGCAACAACTGGGCAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCAATGCCTATCATATTTGGGAAGAGCAGTGGTGACTGTATTGCAACAACTGGGCAAAT 900

Qy    901 GAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATA 960
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATA 960

Qy    961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAAATCAGGACACACGGCACTC 1020
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAAATCAGGACACACGGCACTC 1020

Qy   1021 ACCATAATTCAAAGCACGGGGTCAAGTTCAATGTCATCCCCCAGTGGGCCCAGCCCACAGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCATAATTCAAAGCACGGGGTCAAGTTCAATGTCATCCCCCAGTGGGCCCAGCCCACAGTC 1080

Qy   1081 AACTTCCGGATTCACCCTGGACAGAGTCCTAGAACTCACGAAGAACATT 1140
          ||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACTTCCGGATTCACCCTGGACAGAGTCCTAGAACTCACGAAGAACATT 1140

Qy   1141 GTGGGCTGATAACAGAGTCCAGTTCCAATGTGTTGAGTGCCTTTGACCCCCTCCCCGTCAGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GTGGGCTGATAACAGAGTCCAGTTCCAATGTGTTGAGTGCCTTTGACCCCCTCCCCGTCAGC 1200

Qy   1201 CCTTCTGATGACAAGGCGTACCAGCGTACCAGTCCGGTCTTC 1260
          |||||||||||||||||||||||||||||||||||||||
Db   1201 CCTTCTGATGACAAGGCGTACCAGCGTACCAGTCCGGTCTTC 1260

Qy   1261 CCGGAAGTCAATATTACTGCCCCACGTTACTTCTTCTAATTGGGCAACACAGACAGAGACGGATTCTTT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCGGAAGTCAATATTACTGCCCCACGTTACTTCTTCTAATTGGGCAACACAGACAGAGACGGATTCTTT 1320

Qy   1321 ACAAACCTCACCACTGGCACCTACAGGTTCTACCCCATCATCTACACAGCCTCAAAGACTTC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ACAAACCTCACCACTGGCACCTACAGGTTCTACCCCATCATCTACACAGCCTCAAAGACTTC 1380
```

RESULT 2
US-60-345-785-3
Sequence 3, Application US/60345785
; GENERAL INFORMATION:
; APPLICANT: CURTIS, Rory A.J.
; TITLE OF INVENTION: 58812, A Human Metallopeptidase Family
; TITLE OF INVENTION: Member and Uses Therefor
; FILE REFERENCE: MI2002-003P1(M)
; CURRENT APPLICATION NUMBER: US/60/345,785
; CURRENT FILING DATE: 2002-01-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 1509
;   TYPE: DNA
;   ORGANISM: Homo sapien
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1509)
US-60-345-785-3

Query Match   99.8%;   Score 1505.8;   DB 78;   Length 1509;
Best Local Similarity   99.9%;   Pred. No. 0;
Matches 1507;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

```
Qy      1 ATGGCTCAGCGGTGCGTTGCGTTGCGTCGTGCGTATGCGTATGCGTTATGCTCTAGTTTTCCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTCAGCGGTGCGTTGCGTTGCGTCGTGCGTATGCGTATGCGTTATGCTCTAGTTTTCCCT 60

Qy     61 ACCGTCTCCAGATCGATGGGCCCCGGAGGAGGGGGAGCATCAAAGGGCGTGCGCGGAATCCCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCGTCTCCAGATCGATGGGCCCCGGAGGAGGGGGAGCATCAAAGGGCGTGCGCGGAATCCCT 120

Qy    121 TCTCAGTTCAGCAAAGAGAGAACGCGTCCGGATGAAAGAAGTGCCATCCAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTCAGTTCAGCAAAGAGAGAACGCGTCCGGATGAAAGAAGTGCCATCCAG 180

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACACAGCCCTGGCTGACTTCGGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACACAGCCCTGGCTGACTTCGGA 240

Qy    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTTATCCAGCCATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTTATCCAGCCATGAAGTC 300

Qy    301 GTGGAAGAGTATAGCCACCTGTTCACTCAGTTCTGCTCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAAGAGTATAGCCACCTGTTCACTCAGTTCTGCTCAGCCCTAC 360

Qy    361 CTGCTCGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGCTCGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 420

Qy    421 TTCTCTCGGGTTGGAGCCTGATGCGGTCATCTGGACGACAAGAAC 480
          |||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTCTCTCGGGTTGGAGCCTGATGCGGTCATCTGGACGACAAGAAC 480

Qy    481 TCTGTGAATGGCATTACTGCAGGCCCTGGGAGCTCCTGCTGATCAGGAGGAAGTACATCCCCGA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTGTGAATGGCATTACTGCAGGCCCTGGGAGCTCCTGCTGATCAGGAGGAAGTACATCCCCGA 540
```

```
Qy    541 AGATCTTTCTTCATTTCATCTCGGGCCCATGATGAGGACAGGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGATCTTTCTTCATTTCATCTCGGGCCCATGATGAGGACAGGGGCTCAGAGG 600

Qy    601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACAGGGGGGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACAGGGGGGC 660

Qy    661 TTCATCTTTGGATGATTTCATTCCTTACTTCAAGAAGCCCATCGCCCTTGATTGCAGTCTCA 720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCATCTTTGGATGATTTCATTCCTTACTTCAAGAAGCCCATCGCCCTTGATTGCAGTCTCA 720

Qy    721 GAGAAGGGTTCATTGAACCTAGCTGCGCAGTAAAACATGACTTCAGGCCACTCTTCAGCT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAGAAGGGTTCATTGAACCTAGCTGCGCAGTAAAACATGACTTCAGGCCACTCTTCAGCT 780

Qy    781 CCTCCAAAGGAGGACAAGCCATTGGCATCCTTGCAGCCTCGTCGTGTCAGCCGATTGGAGCAGACA 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCTCCAAAGGAGGACAAGCCATTGGCATCCTTGCAGCCTCGTCGTGTCAGCCGATTGGAGCAGACA 840

Qy    841 CCAATGCCTATCATATTGGAAGCGGGACAGTGTGGACTGTATGTGCAGCAACTGGCAAAT 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCAATGCCTATCATATTGGAAGCGGGACAGTGTGGACTGTATGTGCAGCAACTGGCAAAT 900

Qy    901 GAGTTTCCCTTCCCTGTCAATATAATAATCCTGAGCAACCCATGCTATTTGAACCACTTATA 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 GAGTTTCCCTTCCCTGTCAATATAATAATCCTGAGCAACCCATGCTATTTGAACCACTTATA 960

Qy    961 AGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGGACCACACACGGCACTC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961 AGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGGACCACACACGGCACTC 1020

Qy   1021 ACCATATTCAAAGCAGGGGTCAAGTTCAAAGTCATCCCCCAGTGGCCCCAGGCCCACAGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021 ACCATATTCAAAGCAGGGGTCAAGTTCAAAGTCATCCCCCAGTGGCCCCAGGCCCACAGTC 1080

Qy   1081 AACTTCCGGATTCACCCTGGACACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081 AACTTCCGGATTCACCCTGGACACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1140

Qy   1141 GTGGCTGATAACACAGTCCAGTTCCATGTTGTTGAGTGCCTTTGACCCCTCCCGTCAGC 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141 GTGGCTGATAACACAGTCCAGTTCCATGTTGTTGAGTGCCTTTGACCCCTCCCGTCAGC 1200

Qy   1201 CCTTCTGATGACAAGGGCCTTGGGGTACCAGAGGCTGTCGCCCGACCGTACAGTCCGTTC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201 CCTTCTGATGACAAGGGCCTTGGGGTACCAGAGGCTGTCGCCCGACCGTACAGTCCGTTC 1260

Qy   1261 CCGGAAGTCAATATTACTCGCCCCCAGTTACTTCTATTGGCAACACAGACAGATCGTTCTTT 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261 CCGGAAGTCAATATTACTCGCCCCCAGTTACTTCTATTGGCAACACAGACAGATCGTTCTTT 1320

Qy   1321 ACAAACCTCACCACTGGCATCACAGGTTCTACCCCATCACTACAGCCTGAAGACTTC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321 ACAAACCTCACCACTGGCATCACAGGTTCTACCCCATCACTACAGCCTGAAGACTTC 1380

Qy   1381 AAACGGCATCCAATGGACTCAACGAGAAAATCTCAGTGCCCTTATGACCAAGTGAAA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381 AAACGGCATCCAATGGACTCAACGAGAAAATCTCAGTGCCCTTATGACCAAGTGAAA 1440

Qy   1441 TTCATCTTTGAGTTGATTGAATTCAGAAATCCTGACACAGACCAGACCAGAGACCCACCTTGCAC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441 TTCATCTTTGAGTTGATTGAATTCAGAAATCCTGACACAGACCAGACCAGAGACCCACCTTGCAC 1500

Qy   1501 AAACTGTGA 1509
          |||||||||
Db   1501 AAACTGTGA 1509
```

RESULT 3
US-60-345-785-1

Sequence 1, Application US/6034578S
GENERAL INFORMATION:
APPLICANT: CURTIS, Rory A.J.
TITLE OF INVENTION: 58812, A Human Metallopeptidase Family
TITLE OF INVENTION: Member and Uses Therefor
FILE REFERENCE: MPI2002-003P1(M)
CURRENT APPLICATION NUMBER: US/60/345,785
CURRENT FILING DATE: 2002-01-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 1
LENGTH: 2229
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (53)...(1561)

US-60-345-785-1

Query Match   99.8%; Score 1505.8; DB 78; Length 2229;
Best Local Similarity 99.9%; Pred. No. 0; Mismatches 2; Indels 0; Gaps 0;
Matches 1507; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy      1  ATGGCTCAGCGGTGCGGTTTGCGTTTGCGTCTGCTCGGCCCCTTGGTGGTGGTCGGCCCCCTTGGTGGCTATGCTGCTGCCTGAGTTTTTCCCT   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     53  ATGGCTCAGCGGTGCGGTTTGCGTTTGCGTCTGCTCGGCCCCTTGGTGGTGGTCGGCCCCCTTGGTGGCTATGCTGCTGCCTGAGTTTTTCCCT  112

Qy     61  ACCGTCTCCCAGATCGATGGGGCCGAGGAGCGGGGAGACTCAAAAGGGCGTCGCGAATCCCT  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    113  ACCGTCTCCCAGATCGATGGGGCCGAGGAGCGGGGAGACTCAAAAGGGCGTCGCGAATCCCT  172

Qy    121  TCTCAGTTCAGCAAAAGAGGAACAGCGGTCGCGATGAAGATGCCATCCAG  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    173  TCTCAGTTCAGCAAAAGAGGAACGCGTCGCGATGAAAGGTGCCATCCAG  232

Qy    181  ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAATACTACAGCCCCTGGGCTG  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233  ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAATACTACAGCCCCTGGGCTGGA  292

Qy    241  AAATACAATTCAATAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATTCAGCATGAAGTC  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293  AAATACAATTCAATAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATTCAGCATGAAGTC  352

Qy    301  GTGGAAGAGTATAGCCACCTGTTCACTGTATCCAAGGCTCGGACCCCGGA  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353  GTGGAAGAGTATAGCCACCTGTTCACTATACCAAGGCTCGGACCCCGGA  412

Qy    361  CTGCTGATGGGCTCACTTTGATGTGGTGCCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCA  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413  CTGCTGATGGGCTCACTTTGATGTGGTGCCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCA  472

Qy    421  TTCTCTCGGGTTGGAGCGTGATGGGCGTCATCTATGGTCGGGACGACAAGAAC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473  TTCTCTCGGGTTGGAGCGTGATCGGCGTCATCTATGGTCGGGACGACAAGAAC  532

Qy    481  TCTGTGATGGCAATACTGCCAGCCCTTGGAGCCTTCATCAGGAGAAGTACATCCCCCGA  540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533  TCTGTGATGGCAATACTGCCAGCCCTTGGAGCCTTCATCAGGAGAAGTACATCCCCCGA  592

Qy    541  AGANCTTTCTTCATTTCTCTGGGCCCAAGGAGGACTCATCAGGGGACGCAGGGGGTCAGAGG  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db    593  AGANCTTTCTTCATTTCTCTGGGCCCATGAGGAGGAGCTCATCAGGGGACTCAGGGGGTCAGAGG  652

Qy    601  ATCTCAGCCCCTGCTACAGTCAAGGGGCGTCCAGCTAGCCTTCATTGTTGTGGAGCAGGGGGC  660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653  ATCTCAGCCTGCTACAGTCAAGGGGCGTCCAGCTAGCCTTCATTGTTGACGAGGGGGGGC  712

Qy    661  TTCATCTTGGANGANTTTCAATTCCTAACTTCAAGAAGGCCCATCGCGCCTTGANTGTGGAGCTCTCA  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    713  TTCATCTTGGANGANTTTCATTCCTTAACTTCAAGANGAGCCCATCGCGCCTTGANTGCAGTCTCA  772

Qy    721  GAGAAGGGGTTCCATGAACCTCANTGCTGCAAGTAAACATGACTTCAGGCCACACTTATA  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    773  GAGAAGGGGTTCCATGAACCTCANTGCTCCAAGTAAACATGACTTCAGGCCACACTTATA  832
```

RESULT 4
US-10-014-896-3
Sequence 3, Application US/10014896
GENERAL INFORMATION:
APPLICANT: Yu, Xuanchuan
APPLICANT: Miranda, Marlcar
APPLICANT: Friddle, Carl Johan
TITLE OF INVENTION: Novel Human Proteases and
TITLE OF INVENTION: Polynucleotides Encoding the Same
FILE REFERENCE: LEX-0280-USA
CURRENT APPLICATION NUMBER: US/10/014,896
CURRENT FILING DATE: 2001-12-11
PRIOR APPLICATION NUMBER: US 60/255,567
PRIOR FILING DATE: 2000-12-14
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 3
LENGTH: 1086

```
Db    773  GAGAAGGGGTTCCATGAACCTCANTGAACCTCANTGCTGCAAGTAAACATGACTTCAGGCCACACTTCAGCT  832

Qy    781  CCTCCAAAGGAGGACAAGCAATGGCATCCTTGCCAGCTGCTCCAGCTGGTTGGACGAGCA  840
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    833  CCTCCAAAGGAGGACAAGCAATGGCATCCTTGCCAGCTGCTCCAGCTGGTTGGACGAGCA  892

Qy    841  CCAATGCCTATCATATTGGAAGCGGGACACGTGGTGACTGTATTGCAGCAACTGCAAAT  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    893  CCAATGCCTATCATATTGGAAGCGGGACACGTGGTGACTGTATTGCAGCAACTGCAAAT  952

Qy    901  GAGTTTCCCTTCCCTGTCAAATATAATAATCCTGAGCAACACCCATGGGCTATTTGAACCACACTTATA  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    953  GAGTTTCCCTTCCCTGTCAAATATAATAATCCTGAGCAACACCCATGGGCTATTTGAACCACACTTATA  1012

Qy    961  AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAAATAATCAGGACACCACCACGGCACTC  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1013  AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAAATAATCAGGACACCACCACGGCACTC  1072

Qy   1021  ACCATTCAAACACAGGGGGTCAAGTTCATCATCCCCCAGTGGCCCCCAGGGCCCACAGTC  1080
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1073  ACCATTCAAACACAGGGGGTCAAGTTCATCATCCCCCAGTGGCCCCCAGGGCCCACAGTC  1132

Qy   1081  AACTTCCGGATTCACCCTGGACAGAGCAGTCCTAGAACTCACGAAGGAAGAACATT  1140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1133  AACTTCCGGATTCACCCTGGACAGAGCAGTCCTAGAACTCACGAAGGAAGAACATT  1192

Qy   1141  GTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGGCCTTTGACCCCCCGTCAGC  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1193  GTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGGCCTTTGACCCCCCGTCAGC  1252

Qy   1201  CCTTCTGATGACAAAGGCCTTGGGCTACCAGTCGCCGCCAGACCGTACAGTCCGGTCTTC  1260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1253  CCTTCTGATGACAAAGGCCTTGGGCTACCAGTCGCCGCCAGACCGTACAGTCCGGTCTTC  1312

Qy   1261  CCGGAAGTCAAATTACTGCCCCAGTTACTTCTATTGGCAACACAACAGCACCGGATTCTTT  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1313  CCGGAAGTCAAATTACTGCCCCAGTTACTTCTATTGGCAACACAACAGCACCGGATTCTTT  1372

Qy   1321  ACAAACCTCACCACTGCCATCTACAGTGTCTCACCCCATCTACAACACAGAGCTCGAGAGACTTC  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1373  ACAAACCTCACCACTGCCATCTACAGTGTCTCACCCCATCTACAACACAGAGCTCGAGAGACTTC  1432

Qy   1381  AAACGCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTAGGAGACCAAGTGAAA  1440
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1433  AAACGCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTAGGAGACCAAGTGAAA  1492

Qy   1441  TTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGCAGTTTCTCACCTGCAC  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1493  TTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGCAGTTTCTCACCTGCAC  1552

Qy   1501  AAACTGTGA  1509
           |||||||||
Db   1553  AAACTGTGA  1561
```

TYPE: DNA
ORGANISM: Homo sapiens
US-10-014-896-3

Query Match   67.2%;   Score 1014;   DB 38;   Length 1086;
Best Local Similarity   94.6%;   Pred. No. 2.2e-286;
Matches 1086;   Conservative   0;   Mismatches 0;   Indels 62;   Gaps 1;

```
Qy    1 ATGGCTCAGCGGGTGGGTTGGGTTTGCGTGTGGTGGCCCTGGTTGGCTGCTCCTTAGTTTCCCT   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 ATGGCTCAGCGGGTGGGTTGGGTTTGCGTGTGGTGGCCCTGGTTGGCTGCTCCTTAGTTTCCCT   60

Qy   61 ACCGGTCTCCAGATCGATCGATTGGGCCCGAGGAGGGGGCATCAAAAGGCGTCGGAAATCCCT  120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCGGTCTCCAGATCGATCGATTGGGCCCGAGGAGGGGGCATCAAAAGGCGTCGGAAATCCCT  120

Qy  121 TCTCAGTTCAGCAAAGAGGAACGCGTTCGCCGATGAAAGAGTTGCCATCCAG  180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 CTGCTCAGTTCAGCAAAGAGGAACGCGCTCGCCGATGAAAGAGTTGCCATCCAG  180

Qy  181 ATTCCAACAGTGACTTTTAGCTTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA  240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ATTCCAACAGTGACTTTTAGCTTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA  240

Qy  241 AAATACATTCATAAAGTCTTTCCTACAGTTGTCTCAGCACCACAGCTTTATCCAGCATGAAGTC  300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCACCAGCTTTATCCAGCATGAAGTC  300

Qy  301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAAGGCTCGGACCCCAGCTTGCAGCCCTAC  360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAAGGCTCGGACCCCAGCTTGCAGCCCTAC  360

Qy  361 CTGCTGATGGGCTCACTTTGATGTGGTGGTTGCCTGGCCCGGGAGGAGGTGCCCCCA  420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 CTGCTGATGGGCTCACTTTGATGTGGTGGTTGCCTGGCCCGGGAGGAGGTGCCCCCA  420

Qy  421 TTCTCTCGGGTTGGAGCGGCTGAGTGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAAC  480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 TTCTCTCGGGTTGGAGCGGCTGAGTGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAAC  480

Qy  481 TCTGTGATGGGCAATACTGCCAGGCCTTGGAGCTCCTCGCTCTTGATCAGGAGGTACATCCCCGA  540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 TCTGTGATGGGCAATACTGCCAGGCCTTGGAGCTCCTCGCTCTTGATCAGGAGGTACATCCCCGA  540

Qy  541 AGATCTTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCAGGAGAGGCTCAGAGG  600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 AGATCTTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCAGGAGAGGCTCAGAGG  600

Qy  601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACCGAGGGGGGC  660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACCGAGGGGGGC  660

Qy  661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCTCCCTTGATTGCAGTCTCA  720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCTCCCTTGATTGCAGTCTCA  720

Qy  721 GAGAAGGGTTCCATGAACCTCATCGCTGCAAGTAAACATGACTTCAGCCCACTCTTCAGCT  780
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  721 GAGAAGGGTTCCATGAACCTCATCGCTGCAAGTAAACATGACTTCAGCCCACTCTTCAGCT  780

Qy  781 CCTCCAAAGGAGGAGACAAGCAATTGGCCATCCTGCTCGTGTCAGCTGCAGCTGGAGCCAGACA  840
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  781 CCTCCAAAGGAGACAAGCAATTGGCCATCCTGCTCGTGTCAGCTGCAGCTGGAGCAGACA  840

Qy  841 CCAATGCCTATCATATTTGGAAGCGGGGACAGTGGTGACTGTTATTGCAGCAACTGGCAAT  900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  841 CCAATGCCTATCATATTTGGAAGCGGGGACAGTGGTGACTGTTATTGCAGCAACTGGCAAT  900

Qy  901 GAGTTCCTTCCCTGTCAATATAATCCTGAGCAACCCATCAGGCTGTATTTGAAACCACTTATA  960
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  901 G------------------------------------------------------------  901

Qy  961 AGCAGGTTTATTGGAGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTC 1020
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  961 AGCAGGTTTATTGGAGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTC 1020
```

```
Db  902 ---AGGTTTATTGGAGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTC  958
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy 1021 ACCATATTCAAAAGCAGGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCACAGTC 1080
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  959 ACCATATTCAAAAGCAGGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCACAGTC 1018

Qy 1081 AACTTCCGGAATCACCCTGGACAGCAGAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1019 AACTTCCGGAATCACCCTGGACAGCAGAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1078

Qy 1141 GTGGCTGA 1148
        ||||||||
Db 1079 GTGGCTGA 1086
```

RESULT 5
US-09-648-365-17
; Sequence 17, Application US/09648365
; GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding
; TITLE OF INVENTION: Human Protease-like Homologs
; FILE REFERENCE: 5800-165
; CURRENT APPLICATION NUMBER: US/09/648,365
; CURRENT FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 932
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-648-365-17

Query Match   57.0%;   Score 860.4;   DB 25;   Length 932;
Best Local Similarity   99.9%;   Pred. No. 2.9e-241;
Matches 861;   Conservative   0;   Mismatches 0;   Indels 1;   Gaps 0;

```
Qy  257 TCTTTTCCTACAGTGGTCAGCACCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCC  316
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 TCTTTTCCTACAGTGGTCAGCACCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCC   60

Qy  317 ACCTGTTCACTATCCAAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGGCTCACT  376
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 ACCTGTTCACTATCCAAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGGCTCACT  120

Qy  377 TTGATGTGGTGGTTGCCTGGCCCGGGAGGAGGTGCCCCCATTCTCTCGGGTTGGAGCGGCTGGAGC  436
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 TTGATGTGGTGGTTGCCTGGCCCGGGAGGAGGTGCCCCCATTCTCTCGGGTTGGAGC  180

Qy  437 GTGATGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAACTCTGTGATGGGCATTAC  496
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 GTGATGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAACTCTGTGATGGGCATTAC  240

Qy  497 TGCAGGCCTTGGAGCTCCTCGCTCTTGATCAGGAGGTACATCCCCGAAGATCTTTTCTTCATTT  556
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 TGCAGGCCTTGGAGCTCCTCGCTCTTGATCAGGAGGTACATCCCCGAAGATCTTTTCTTCATT  300

Qy  557 CTCTGGGCCCATGATGAGGAGTCATCAGGAGAGGCTCAGAGGATCTCAGCCCTGCTAC  616
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 CTCTGGGCCCATGATGAGGAGTCATCAGGAGAGGCTCAGAGGATCTCAGCCCTGCTAC  360

Qy  617 AGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACCGAGGGGGGCTTCATCTTGGATGATT  676
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361 AGTCAAGGGCGTCCAGCTAGCCTTCATTGTGGACCGAGGGGGGCTTCATCTTGGATGATT  420

Qy  677 TCATTCCTAACTTCAAGAAGCCCATCGCCTCCCTTGATTGCAGTCTCAGAGAAGGGGTTCCATGA  736
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  421 TCATTCCTAACTTCAAGAAGCCCATCGCCTCCCTTGATTGCAGTCTCAGAGAAGGGTTCCATGA  480

Qy  737 ACCTCATCGCTGCAAGTAAACATGACTTCAGCCCACTCTTCAGCTCCTCCAAAAGGAGGACA  796
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  481 ACCTCATCGCTGCAAGTAAACATGACTTCAGCCCACTCTTCAGCTCCTCCAAAAGGAGGACA  540
```

RESULT 6
US-10-259-776-17
Sequence 17, Application US/10259776
GENERAL INFORMATION:
; APPLICANT: Robison, Keith E.
; TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding
; TITLE OF INVENTION: Human Protease-Like Homologs
; FILE REFERENCE: 5800-165
; CURRENT APPLICATION NUMBER: US/10/259,776
; CURRENT FILING DATE: 2002-09-27
; PRIOR APPLICATION NUMBER: US/09/648,365
; PRIOR FILING DATE: 2000-08-25
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 932
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-259-776-17

Query Match   57.0%;
Best Local Similarity   99.9%;   Score 860.4;   DB 42;   Length 932;
Matches   861;   Conservative   0;   Pred. No. 2.9e-241;   Mismatches   1;
   Indels   0;   Gaps   0;

```
Qy    797 GCATTGGCATCCTTGCAGCGTGCGTGTCAGCGGATGGGAGGCAGAGACCAATGCCTATCATAT 856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCATTGGCATCCTTGCAGCGTGCGTGTCAGCGGATGGGAGGCAGAGACCAATGCCTATCATAT 600

Qy    857 TTGGAAGCGGGGACAGTGGTGTTGTATTGCAGCAACTGCAAATGAGTTTCCTTCCCTG 916
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGGAAGCGGGGACAGTGGTGTTGTATTGCAGCAACTGCAAATGAGTTTCCTTCCCTG 660

Qy    917 TCAATATAATCCTGAGCAACCATGGCTATTTGAACCACTAGTAAAGCAGGTTTATGGAGA 976
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCAATATAATCCTGAGCAACCATGGCTATTTGAACCACTAGTAAAGCAGGTTTATGGAGA 720

Qy    977 GAAATCCCTTAACCAATGCAATAATCAGGACCACACCACGGGCCACTCACCATATTCAAAAGCAG 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAAATCCCTTAACCAATGCAATAATCAGGACCACACCACGGGCCACTCACCATATTCAAAAGCAG 780

Qy   1037 GGGTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGGCCACAGTCAAACTTCCGGATTCACC 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGGCCACAGTCAAACTTCCGGATTCACC 840

Qy   1097 CTGGACAGACAGTCCAAGAGGT 1118
          ||||||||||||||||||||||
Db    841 CTGGACAGACAGTCCAAGAGGT 862
```

RESULT 7
US-09-666-026-75
Sequence 75, Application US/09666026
GENERAL INFORMATION:
; APPLICANT: Hodgson, David M.
; APPLICANT: Lincoln, Stephen E.
; APPLICANT: Russo, Frank D.
; APPLICANT: Spiro, Peter A.
; APPLICANT: Banville, Steve C.
; APPLICANT: Bratcher, Shawn R.
; APPLICANT: Dufour, Gerard E.
; APPLICANT: Cohen, Howard J.
; APPLICANT: Rosen, Bruce H.
; APPLICANT: Shah, Purvi
; APPLICANT: Chalup, Michael S.
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Jones, Anissa L.
; APPLICANT: Yu, Jimmy Y.
; APPLICANT: Greenawalt, Lila B.
; APPLICANT: Panzer, Scott R.
; APPLICANT: Roseberry, Ann M.
; APPLICANT: Wright, Rachel J.
; APPLICANT: Chen, Wensheng
; APPLICANT: Liu, Tommy F.
; APPLICANT: Yap, Pierre E.
; APPLICANT: Stockdreher, Theresa K.
; APPLICANT: Amshey, Stefan
; APPLICANT: Fong, Willy T.
; TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES
; FILE REFERENCE: PF-0073 US
; CURRENT APPLICATION NUMBER: US/09/666,026
; CURRENT FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 60/156,296
; PRIOR FILING DATE: 1999-09-24
; PRIOR APPLICATION NUMBER: 60/156,565

```
Qy    617 AGTCAAGGGGCGGCTGCCAGGCTAGCCTTCATTGTGGACGGAGGGGGGCGTTCATCTTGGATGATT 676
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 AGTCAAGGGGCGGCTCCAACTAGCCTTCATTGTGGACGGAGGGGGGCTTCATCTTGGATGATT 420

Qy    677 TCATTCCTAACTTCAAGAAAGCCCCATCGCCTTGATTGCAGTCTTCAGAGAAGGGTTCCATGA 736
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TCATTCCTAACTTCAAGAAAGCCCCATCGCCTTGATTGCAGTCTTCAGAGAAGGGTTCCATGA 480

Qy    737 ACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGTCTCCAAAAGGAGACAA 796
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 ACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGTCTCCAAAAGGAGACAA 540

Qy    797 GCATTGGCATCCTTGCAGCGTGCGTCAGCGGATGGGAGGCAGAGACCAATGCCTATCATAT 856
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCATTGGCATCCTTGCAGCGTGCGTCAGCGGATGGGAGGCAGAGACCAATGCCTATCATAT 600

Qy    857 TTGGAAGCGGGGACAGTGGTGTTGTATTGCAGCAACTGCAAATGAGTTTCCTTCCCTG 916
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 TTGGAAGCGGGGACAGTGGTGTTGTATTGCAGCAACTGCAAATGAGTTTCCTTCCCTG 660

Qy    917 TCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTAGTAAAGCAGGTTTATGGAGA 976
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTAGTAAAGCAGGTTTATGGAGA 720

Qy    977 GAAATCCCTTAACCAATGCAATAATCAGGACCACACCACGGGCCACTCACCATATTCAAAAGCAG 1036
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAAATCCCTTAACCAATGCAATAATCAGGACCACACCACGGGCCACTCACCATATTCAAAAGCAG 780

Qy   1037 GGGTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGGCCACAGTCAAACTTCCGGATTCACC 1096
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GGGTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGGCCACAGTCAAACTTCCGGATTCACC 840

Qy   1097 CTGGACAGACAGTCCAAGAGGT 1118
          ||||||||||||||||||||||
Db    841 CTGGACAGACAGTCCAAGAGGT 862
```

```
Qy    257 TCTTTCCTACAGTGGTGTCAGCACCAGCTTTATCCACATGAAGTCGTGGAAGAGTATAGCC 316
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 TCTTTCCTACAGTGGTGTCAGCACCAGCTTTATCCACATGAAGTCGTGGAAGAGTATAGCC 60

Qy    317 ACCTGTTCACTATCCAAGGCTCGGACCCAGCTTGCAGCCCTACCTGCTGATGGCTCACT 376
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCTGTTCACTATCCAAGGCTCGGACCCAGCTTGCAGCCCTACCTGCTGATGGCTCACT 120

Qy    377 TTGATGTGGTGCCTGCCCCTGAAGGAAGAGGCTGGGAGGTGGGACC 436
          ||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TTGATGTGGTGCCTGCCCCTGAAGGAAGAGGCTGGGAGGTGGGACC 180

Qy    437 GTGATGGCGTCATCTATGGTGGGCACGACGAAGAACTCTGTGATGTGGCATAC 496
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 GTGATGGCGTCATCTATGGTGGGCACGACGAAGAACTCTGTGATGTGGCATAC 240

Qy    497 TGCAGGCCTTGGAGCTCCTGCTCCTGATCAGGAAGAAGTACATCCCCCGAAGAATCTTTCTTCATTT 556
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 TGCAGGCCTTGGAGCTCCTGCTCCTGATCAGGAAGAAGTACATCCCCGAAGAATCTTTCTTCATTT 300

Qy    557 CTCTGGGCCATGATGAGGAGTCATCAGGACAGGGGCTCAGAGGAATCTCAGCCCTGCTAC 616
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CTCTGGGCCATGATGAGGAGTCATCAGGACAGGGGCTCAGAGGAATCTCAGCCCTGCTAC 360
```

PRIOR APPLICATION FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: 60/156,624
PRIOR APPLICATION FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: 60/156,625
PRIOR APPLICATION FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: 60/167,520
PRIOR APPLICATION FILING DATE: 1999-11-24
PRIOR APPLICATION NUMBER: 60/168,197
PRIOR APPLICATION FILING DATE: 1999-11-30
PRIOR APPLICATION NUMBER: 60/168,614
PRIOR APPLICATION FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/168,611
PRIOR APPLICATION FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: 60/168,613
PRIOR APPLICATION FILING DATE: 1999-12-02

NUMBER OF SEQ ID NOS: 258
SOFTWARE: PERL Program

SEQ ID NO 75
LENGTH: 1271
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 369881.2.oct
NAME/KEY: unsure
LOCATION: 32
OTHER INFORMATION: a, t, c, g, or other
US-09-666-026-75

Query Match   44.6%;   Score 673.4; DB 26; Length 1271;
Best Local Similarity   99.7%;   Pred. No. 3.2e-186;
Matches 674; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   834 GCAGACACCAAATGCCTATCATATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACT 893
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCAGACACCAAATGCCTATCATATATTTGGAAGCNGGACAGTGGTGACAGTATTGCAGCAACT 60

Qy   894 GGCAAATGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCATGGGCTATTTGAACC 953
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCAAATGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCATGGGCTATTTGAACC 120

Qy   954 ACTTATAAGCACAGGTTTATTGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCAC 1013
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTTATAAGCACAGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCAC 180

Qy   1014 GGCACTCACCATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGC 1073
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCACTCACCATATTCAAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGC 240

Qy   1074 CACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCTTAGAACTCACGAA 1133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCTTAGAACTCACGAA 300

Qy   1134 GAACATTGTGGGCTGGATAAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTGACCCCCTCC 1193
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 GAACATTGTGGGCTGCATAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTTGACCCCCTCC 360

Qy   1194 CGTCAGCCCTTCTGATGACAAGGCCTTGGGCTACCAGCGTGCTCCGCGAGACCGTACAGTC 1253
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 CGTCAGCCCTTCTGATGACAAGGCCTTGGGCTACCAGCTGCTCCGCGAGACCGTACAGTC 420

Qy   1254 CGTCTTCCGGGAGTCAATATTACTGCCCCAGTTACTTCTATTGGCAACACAGACCACCAC 1313
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 CGTCTTCCGGGAGTCAATATTACTGCCCCAGTTACTTCTATTGGCAACACAGACCACCACCG 480

Qy   1314 ATTCTTTACAAAACCTCACACTGGCATCTTACCACCGGTTCTTACCCCAATCTACATAACAGCCTGA 1373
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ATTCTTTACAAAACCTCACCACTGGCATCTTCACCACCGGTTCTTACCCCAATCTACATAACAGCCTGA 540

Qy   1374 AGACTTCAAAACGCATCCATGGAGTCAAACGAGAAAAATCTCAGTCCATGTATGAGACCCA 1433
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 AGACTTCAAAACGCATCCATGGAGTCAACGAGAAAAATCTCAGTCCATGTATGAGACCCA 600

RESULT 8
US-60-156-625-265
Sequence 265, Application US/60156625
GENERAL INFORMATION:
APPLICANT: Hodgson, David M.
APPLICANT: Lincoln, Stephen E.
APPLICANT: Russo, Frank D.
APPLICANT: Spiro, Peter A.
APPLICANT: Banville, Steve C.
APPLICANT: Bratcher, Shawn R.
APPLICANT: Dufour, Gerard E.
APPLICANT: Cohen, Howard J.
APPLICANT: Rosen, Bruce
APPLICANT: Shah, Purvi
APPLICANT: Chalup, Michael S.
APPLICANT: Hillman, Jennifer L.
APPLICANT: Jones, Anissa L.
APPLICANT: Yu, Jimmy Y.
APPLICANT: Greenawalt, Lila B.
APPLICANT: Panzer, Scott R.
APPLICANT: Roseberry, Ann M.
APPLICANT: Wright, Rachel J.
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-0088 P
CURRENT APPLICATION NUMBER: US/60/156,625
CURRENT FILING DATE: 1999-09-28
NUMBER OF SEQ ID NOS: 4230
SOFTWARE: PERL Program

SEQ ID NO 265
LENGTH: 1271
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: unsure
LOCATION: 32
OTHER INFORMATION: a or g or c or t, unknown, or other
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte template ID No: 369881.2
US-60-156-625-265

Query Match   44.6%;   Score 673.4; DB 59; Length 1271;
Best Local Similarity   99.7%;   Pred. No. 3.2e-186;
Matches 674; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   834 GCAGACACCAAATGCCTATCATATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACT 893
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GCAGACACCAAATGCCTATCATATATTTGGAAGCNGGACAGTGGTGACAGTATTGCAGCAACT 60

Qy   894 GGCAAATGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCATGGGCTATTTGAACC 953
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 GGCAAATGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCATGGGCTATTTGAACC 120

Qy   954 ACTTATAAGCACAGGTTTATTGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCAC 1013
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACTTATAAGCACAGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCAC 180

Qy   1014 GGCACTCACCATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGC 1073
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GGCACTCACCATATTCAAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGC 240

Qy   1074 CACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCTTAGAACTCACGAA 1133
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 CACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCTTAGAACTCACGAA 300

Qy   1434 AGTGAAATTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGGAGGACCCAGTTTCTCA 1493
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 AGTGAAATTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGGAGGACCCAGTTTCTCA 660

Qy   1494 CCTGCACAAACTGTGA 1509
          ||||||||||||||||
Db   661 CCTGCACAAACTGTGA 676

```
Qy  1134 GAACATTGTGGCTGATAAACAGAGAGTCCAGTCATGTGTGAGTGGTGGTGCCCCTCC 1193
          |||||||||||||||||||||||| ||||||| |||||| ||||||  | ||||||
Db   301 GAACATTGTGGCTGATAAACAGAGTCCAGTCCATGTGTTGAGTGGTTTGACCCCTCCC 360

Qy  1194 CGTCAGCCCTTCTGATGACAAAGGCCTTGGGGCTACCAGTGTCCATGTCCGGTACAGTC 1253
          ||||||||||||||||||||||||||||||| ||||||| |||| | |||||||||||
Db   361 CGTCAGCCCTTCTGATGACAAAGGCCTTGGGCTACCAGCGTCCTCCGCCAGACCGGTACAGTC 420

Qy  1254 CGTCTCCCGGAAGTCAATATTATTCTGCCCCAGTAGTACTCTATTGGCAACACAGACGCCG 1313
          |||||||||||||||||| || ||| ||||||| | || || | ||||| |||||
Db   421 CGTCTCCCGGAAGTCAATAATTACTCCCCAGTCAATATTACTCCCAGTCAGAAAACTACACCCCAAA 480

Qy  1314 ATTCTTTTACAAACCTCACCCACTCACCACTGGGAATCTACAGGTTCTCACCCANTCTACAGTACATACAGCCTGA 1373
          ||||||| ||||||||||| |||   ||||||||||||||||||||||||   |  | |||| ||| |||||||||
Db   481 ATTCTTTACAAACCTCACCACTGGGAATCTACAGGTTCTCACTCTTCTTATTGGCACACAGACAGCCTGA 540

Qy  1374 AGACTTCAAAGCCATCCATGAGTCAAACGACAGAAAATCTCAGTCATCCACATACAGCCCA 1433
          ||||||||||||||||||| ||||||||||||||||||||||| |||   |||| ||||||
Db   541 AGACTTCAAAGCCATCCATGGAGTCAAACGACAGAAAATCTCAGTCGTCAAGCCTTATGAGACCCA 600

Qy  1434 AGTGAAAATTCATCTTTGAGTGTGATTCAGAATGTCACACAGACCAGGAGCCAGTTTCTCA 1493
          |||||||||||||||||||| |||| |||||   |||||||||||||||||||||||||||
Db   601 AGTGAAAATTCATCTTTGAGTTGTGACACAGAACCAGGAGCCAGTTTCTCA 660

Qy  1494 CCTGCACAAACTGTGA 1509
          ||||||||||||||||
Db   661 CCTGCACAAACTGTGA 676
```

RESULT 9
US-60-360-207-413
Sequence 413, Application US/60360207

Query Match   39.5%; Score 596.8; DB 80; Length 909;
Best Local Similarity   80.5%; Pred. No. 8.9e-164;
Matches 729; Conservative 0; Mismatches 162; Indels 15; Gaps 2;

```
Qy    1 ATGGCTCAGCGCGGTGCCGTTGCCGTGGTGCCCGTATCGCTGCGCTCCTCCTAGTTTCCT 60
         | ||||| ||  ||  |  | || |||||||||   |||  |   ||||||||||  ||
Db   16 ATGGCTGAGCTACTTGCCTAGCTGTGCCGGCCAGCTGTCGTTCTTGTTTGTTTGTTTCGCT 75

Qy   61 ACGGTCTCTCCAGATCGATGGGCCCGAGGAGCGCGGGAGCATCAAAGGCGTGCGAATCCCT 120
         ||  |  | || ||||  |  |||||||||||||  ||||||||||||||| |||||||
Db   76 ACGGTCTCCGGATCCACTGGGCCTAGAAGCAGGAGAAAATCGGGAGGGGGGAATCCCT 135

Qy  121 TCTCAGTTCAGCAAAAGAGGACGACGCGCTGCGCGCTGAAAAGTGCCATCCAG 180
         || |||||||||  || || || |||| |  ||  ||| ||||| |||||||||
Db  136 TCCAGTTCAGCGAGGAGGGGTGTCGCTATAAAGAGAGGGCCTGAAAAGGTGCCATCCAG 195

Qy  181 ATTCCAACAGTCGACTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGGTGAGTTCGGA 240
         ||||| ||||| | ||||  | |  |  ||||| |  |||||||| | | |||||| |||
Db  196 ATTCCCAACAGTGTCTTTTCAGCCACACACCACAGCCCTTGCTGAGTTTGGA 255

Qy  241 AAATACATTCATAAAGTCTTCCTACAGTGGTCTACCCAGCTTTATCAGCCATGAAGTC 300
         |     |  ||||  | || |||||| |   | ||| |||||  ||||||||||||||
Db  256 --------AGTGTTCCACAGCACCCCTGTCCAACATCAAGTC 303

Qy  301 GTGGAAGTATAGCCACCTGTTCACTATCCAAGGCTCAGCACCCCAGTTTGCAGCCCTAC 360
         ||||  |||||||||||||||||| ||||||  | |||| |||||||||||||||||||
Db  304 GTGGCAAAGTATAGCCACCTGTTCACCATCCAAAGGCCTCAGCCCCAGTTTGCAGCCCTAC 363
```

RESULT 10
PCT-US02-09921-430
Sequence 430, Application PC/TUS0209921

APPLICANT: DAFFO, Abel
APPLICANT: JONES, Anissa L.
APPLICANT: TRAN, Alanna-Phung B.
APPLICANT: DAHL, Christopher R.
APPLICANT: GIETZEN, Darryl
APPLICANT: CHINN, Joyce
APPLICANT: DUFOUR, Gerard E.
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: YU, Jimmy Y.
APPLICANT: TUASON, Olivia
APPLICANT: YAP, Pierre E.
APPLICANT: AMSHEY, Stefan R.
APPLICANT: DAUGHERTY, Sean C.
APPLICANT: DAM, Tam C.
APPLICANT: LIU, Tommy F.
APPLICANT: NGUYEN, Duy-Viet An
APPLICANT: KLEEFELD, Yael
APPLICANT: GERSTIN JR., Edward H.
APPLICANT: PERALTA, Careyna H.
APPLICANT: DAVID, Marie H.
APPLICANT: LEWIS, Samantha A.
APPLICANT: CHEN, Alice J.
APPLICANT: PANZER, Scott R.
APPLICANT: HARRIS, Bernard
APPLICANT: FLORES, Vincent
APPLICANT: MARWAHA, Rakesh

```
Qy  361 CTGCTGATGGCTCACTTGATGTGTGGTGCGCCCCTGCCCCCTGAAGAAAGAGGCTGGGAGGTGCCCCCA 420
         ||||||||||||||| |||||| ||||   | |||||||||| |||||| |   |||| ||||||||
Db  364 ATGCTGATGGCTCACATTGATGTGGTGGTTTCCTGCCCCCGAAGAAGAGANGGGGGTGCCCCG 423

Qy  421 TTCTCTGGGTTGGAGCGTGATGGCGTCATCTAATGGTGGGGGGCACACTGGACGAAGGAAC 480
         |||||||||| ||||| ||| ||||||||| |||||||||||||||||||||| ||||
Db  424 TTCTCAGGCCGGAAGCGCAATAGGGCTTCATCTATAGTGGGGGTGCCTGGACAACAAAAAC 483

Qy  481 TCTGTGATGGCATTACTGCAGGCCTGGGAGCTCGCTGCCTGACTCAGGAAGTACATCCCGGA 540
         |  ||||| |||||||  |||  ||||||||| ||||||| || ||| ||||||||| |||
Db  484 TCTGTGATGGGCGAATCCTGCATGCTTTGGACGTCCTTTGGACAGGTGTCCCGGAATCCCC 543

Qy  541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTCATCAGG---ACAGGGGTCAG 597
         |||||||| ||||| |||| ||||||||| |||| ||||||||    |||||||||||
Db  544 AGATCTTTCTTCATTTGCTTTGGGCCATGATGAGGAGTCTTGTTGATCAGAAAACTACAGCCCAG 603

Qy  598 AGGATCTCAGCCCTGCTCTCAGTCAAGGGGCGTCCACCTACCTTCATTGTTGTGGACGAGGGG 657
         |||| ||||||  || || |||| |||||| ||| ||||| |||| |||| ||||||||||
Db  604 AAGATCTCAGACACTCTTACAGCCAAGGGGGTTCCACGCCTTCCTTGTTTGGACAGAGGG 663

Qy  658 GGCTTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGAGCCCATCGCCCTTGAGTGGCAGTC 717
         ||||||||||||||||||||| |||||   || |||||||| || || ||||||||||||||
Db  664 AGCTTTACCTTGGAAGGCTTCATTCCAAAACCTCGAGAAGAGCCAGTTGCCATCAGATTTCAGTC 723

Qy  718 TCAGAGAAGGGGTTCCATGAAACCTCATCGCTGCAAGTAAACATGACAGCTCTCA 777
         | ||||| || |||||| |||||||||| ||||||||| |||||| | ||||||||
Db  724 ACTGAGAAGGGGTGCCCCTTGACCTCATCGTCTGCAAAGTAAAACATGACCCAGGCCCACTCTTCA 783

Qy  778 GCTCCTCCAAAGGAGACAAGCAATGGCATCCTTGCAGCTGCGTGCCGATTGGAGCAG 837
         ||||||||||||||||||||||||||||||||| |||| ||   |||| ||||||||
Db  784 GCTCCCCAAAGGAGACAAGCAATGGCGATTCTTCTCGCCGCCGACTGGAGCAG 843

Qy  838 ACACCAATGCCTATCATATTTGGAAGGGGACAGTGGTGACTGTATTGCAACTGGCA 897
         |||||||| | |||| |||| ||||| |||||||| |||| ||||| |||||||||
Db  844 ACACCAATGCCGAATATGTTTGGAGGAGGGGCCCATTGAGAGAACAATGAAGCTACTGGCA 903

Qy  898 AATGAG 903
         ||||||
Db  904 AATGAG 909

Qy  904 AATGAG 909
```

APPLICANT: LO, Audrey
APPLICANT: LAN, Ruth Y.
APPLICANT: URASHIMA, Michael
TITLE OF INVENTION: SECRETORY MOLECULES
FILE REFERENCE: PT-1232 PCT
CURRENT APPLICATION NUMBER: PCT/US02/09921
CURRENT APPLICATION FILING DATE: 2002-03-27
PRIOR APPLICATION NUMBER: 60/280,067; 60/280,068; 60/291,280; 60/291,849;
PRIOR APPLICATION NUMBER: 60/291,829; 60/299,428; 60/300,001; 60/299,776
PRIOR FILING DATE: 2001-03-29; 2001-03-29; 2001-05-16; 2001-05-17;
2001-05-17; 2001-06-19; 2001-06-20; 2001-06-20
NUMBER OF SEQ ID NOS: 1146
SOFTWARE: PERL Program
SEQ ID NO 430
LENGTH: 1307
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: LG:369881.5:2001JUN22
PCT-US02-09921-430

Query Match   38.6%; Score 582; DB 1; Length 1307;
Best Local Similarity   89.0%; Pred. No. 2.4e-159;
Matches 687; Conservative   0; Mismatches   0; Indels 85; Gaps   2;

(sequence alignment blocks follow)

RESULT 11
US-09-648-365-18
Sequence 18, Application US/09648365
GENERAL INFORMATION:
APPLICANT: Robison, Keith E.
TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding
TITLE OF INVENTION: Human Protease-like Homologs
FILE REFERENCE: 5800-165
CURRENT APPLICATION NUMBER: US/09/648,365
CURRENT APPLICATION FILING DATE: 2000-08-25
NUMBER OF SEQ ID NOS: 41
SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 18
LENGTH: 722
TYPE: DNA
ORGANISM: Homo sapiens
US-09-648-365-18

Query Match   38.0%; Score 572.8; DB 25; Length 722;
Best Local Similarity   98.8%; Pred. No. 9e-157;
Matches 577; Conservative   0; Mismatches   7; Indels   0; Gaps   0;

RESULT 12
US-10-259-776-18
Sequence 18, Application US/10259776
GENERAL INFORMATION:

```
;   APPLICANT: Robison, Keith E.
;   TITLE OF INVENTION: Novel Nucleic Acid Molecules Encoding
;   TITLE OF INVENTION: Human Protease-like Homologs
;   FILE REFERENCE: 5800-165
;   CURRENT APPLICATION NUMBER: US/10/259,776
;   CURRENT FILING DATE: 2002-09-27
;   PRIOR APPLICATION NUMBER: US/09/648,365
;   PRIOR FILING DATE: 2000-08-25
;   NUMBER OF SEQ ID NOS: 41
;   SOFTWARE: FastSEQ for Windows Version 3.0
;
;   SEQ ID NO 18
;   LENGTH: 722
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-259-776-18

Query Match       38.0%;  Score 572.8;  DB 42;  Length 722;
Best Local Similarity    98.8%;  Pred. No. 9e-157;
Matches 577;  Conservative    0;  Mismatches    7;  Indels    0;  Gaps    0;

Qy   257 TCTTTCCTACACGTGGTCAGCACCAGCTTTATCCAGCATGAAGTCGTTGGAAGAGTATAGCC 316
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 TCTTTCCTACACGTGGTCAGCACCAGCTTTATCCAGCATGAAGTCGTTGGAAGAGTATAGCC 60

Qy   317 ACCTGTTCACTATCCAAGGCGTCGGGACCCCACCAGCTTGCAGCCCTACCTGATGCTCACT 376
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 ACCTGTTCACTATCCAAGGCGTCGGGACCCCACCAGCTTGCAGCCCTACCTGATGCTCACT 120

Qy   377 TTGATGTGGTTGGCTGCCTGCCCCTGAAGAGGCTGGGAGGTGCCCCCATTCTCTCGGTGGAGC 436
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 TTGATGTGGTTGGCTGCCTGCCCCTGAAGAGGCTGGGAGGTGCCCCCATTCTCTCGGTTGGGAGC 180

Qy   437 GTGATGGCGTCATCTATGGGTCGGGGGACACTGGACAGCAAGAACTCTGTGAGTGGCATTAC 496
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 GTGATGGCATCATCTATGGTCGGGGGACACACTGGACAGCAAGAACTCTGTGATGGCATTAC 240

Qy   497 TGCAGGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGGAAGATCTTTCTTCATTT 556
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 TGCAGGGCCTTGAGCTCCTGCTGATCAGGAAGTACATCCCCGGAAGATCTTTCTTCATTT 300

Qy   557 CTCTGGGGCCATGAGAGGAGGAGTCATCAGGAGGACAGGGGCTCAGAGGATCTCAGCCTAC 616
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 CTCTGGGGCCATGATGAGGAGTCATCAGGACAGGGGCTCAGAGGATCTCAGCCCCTGCTAC 360

Qy   617 AGTCAAGGGCGGTCCCACCTAGCCTTCATTGTGGACAGCAGGGGGCTTCATTGTGGATGATT 676
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 AGTCAAGGGCGGTCCCAGCTAGCCTTCATTGTGGACAGCAGGGGGCTTCATTGTGGATGATT 420

Qy   677 TCAATCCTAACTTCAAGAAGTAAACANTGAACTTCAGGCTCCTCCTCCAAAGGAGGACAA 736
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 TCAATCCTAACTTCAAGAAGCCCATCGCCCTGATTGCAGTCTCAGAGAAGGGTTCCATGA 480

Qy   737 ACCTCATGTCGAAGTAAACANTGAACATTCAGGCCCACTCCTCCTCCAAAGGAGGACA 796
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 ACCTTCATCGTGCAAGTAAACATGAACATTCAGGCCCACTCCTCCTCCAAAGGAGGACAA 540

Qy   797 GCAATGGCCATCCTTGCAGCTGCTGGTCCAGCCGAATTGGAGCAGACA 840
         ||||||||||||||||||||||||||||||| ||
Db   541 GCAATGGCCATCCTTGCAGCTGCTGTGTCAGCCGGTTTATGGAGGAGA 584
```

RESULT 13
US-60-213-362-6769/c
; Sequence 6769, Application US/60213362
; GENERAL INFORMATION:
;   APPLICANT: Morris, MacDonald
;   APPLICANT: Lal, Preeti
;   APPLICANT: Diep, Dinh
;   TITLE OF INVENTION: Method for the Identification of Sequence Polymorphisms Using
;   TITLE OF INVENTION: Polynucleotide Sequence Databases, and Single Nucleotide Polymor
;   FILE REFERENCE: GX-0016 P
;   CURRENT APPLICATION NUMBER: US/60/213,362

```
;   CURRENT FILING DATE: 2000-06-21
;   NUMBER OF SEQ ID NOS: 8429
;   SOFTWARE: PERL Program
;
;   SEQ ID NO 6769
;   LENGTH: 649
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No: 423998.2
US-60-213-362-6769

Query Match       31.4%;  Score 473.4;  DB 65;  Length 649;
Best Local Similarity    98.6%;  Pred. No. 1.4e-127;
Matches 488;  Conservative    0;  Mismatches    6;  Indels    1;  Gaps    1;

Qy     1 ATGGCTCAGCGCGGTGCGGTTGCGTGCGTGCGCCCCTGGGTGGCCTARGCTGCTCCTAGTTTTCCT 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   562 ATGGCTCAGCGCGGTGCGGTTGCGTGCGTGCGCCCCTGGGTGGCCTAGCTGCTCCTAGTTTTCCT 503

Qy    61 ACCGTCTCTCCAGATCGATGGGCCCGAGCGAGGAGGAGGCATCAAAAGGCCGTGCGGCCATCCAG 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   502 ACCGTCTCCAGATCGATGGGCCCGAGGAGGAGGCGGGGAGTATCAAAAGGCGTGCGGAATCCCT 443

Qy   121 TCTCAGTTCAGCAAAGAGGAGAACGAACGCGTCGGATGAAAGAGAGAGGCCGGGTGGCCATCCAG 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   442 TCTCAGTTCAGCAAAGAGGAACCGCGTCCGATGAAAGAGAGAGGCACTGAAAGGTGGCCATCCAG 383

Qy   181 ATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAAATACTACAGCCCTGGGTGGCTGAGTTCGGA 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   382 ATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAAATACTACAGCCCTGGGTGGCTGAGTTCGGA 323

Qy   241 AAAATACATTCATAAAAGTCTTTCCTACACGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   322 AAATACATTCGTAAAAGTCTTTCCTACACGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 263

Qy   301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCTCGGGACCCGACCCCAGCCTAC 360
         ||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   262 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGGACCCCA-CTTGCAGCCCTAC 204

Qy   361 CTGCTGATGGCTCACTTTGATGATGGTGGTGCCGGGAGGTGGGAGGTGCCCCA 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   203 CTGCTGATGGCTCACTTTGATGTGGTGCCCCTGAAGAGAGGCTGGGAGGTGCCCCA 144

Qy   421 TTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTGCGGACGACAAGAAC 480
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   143 TTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGACGACAAGAAC 84

Qy   481 TCTGTGATGGCATTA 495
         ||||||||||||| |
Db    83 TCTGTGATGGTCTGA 69
```

RESULT 14
US-60-250-830-1230/c
; Sequence 1230, Application US/60250830
; GENERAL INFORMATION:
;   APPLICANT: Morris, MacDonald
;   APPLICANT: Lal, Preeti
;   APPLICANT: Diep, Dinh
;   TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
;   TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
;   TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
;   FILE REFERENCE: GX-0020 P
;   CURRENT APPLICATION NUMBER: US/60/250,830
;   CURRENT FILING DATE: 2000-11-04
;   NUMBER OF SEQ ID NOS: 3246
;   SOFTWARE: PERL Program
;
;   SEQ ID NO 1230
;   LENGTH: 649
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:

```
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No: 423998.2
US-250-830-1230

Query Match        31.4%;    Score 473.4;   DB 69;   Length 649;
Best Local Similarity   98.6%;   Pred. No. 1.4e-127;
Matches   488;   Conservative    0;   Mismatches   6;   Indels   1;   Gaps   1;

Qy      1 ATGGCTCAGCGGTGCGTTTGCGTGCTGCCGCCGGGCAGCATCAAAAGGGCGTCGCCCATCCAG 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    562 ATGGCTCAGCGGTGCGTTTGCGTGCTGCCGCCGGGCAGCATCAAAAGGGCGTCGCCCATCCAG 503

Qy    121 TCTCAGTTCAGCAAAAGAGAGAAACGCGTCGGGATGAAAAGAGGCGCTGGAAAAGGTGCCATCCAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    502 ACCGTCTCCAGATCGATGGGCCCCGAGGAGCGGGGAGTATCAAAAGGGCGTCGCGAAATCCCT 443

Qy     61 ACCGTCTCCAGATCGATGGGCCCCGAGGAGCGGGGAGCATCAAAAGGGCGTCGCGAAATCCCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    442 TCTCAGTTCAGCAAAAGAGAGAAACGCGTCGGGATGAAAAGAGGCGCTGGAAAAGGTGCCATCCAG 383
```

Search completed: July 1, 2003, 08:49:56
Job time : 3689 secs

```
RESULT 15
US-60-323-966-1230/c
Sequence 1230, Application US/60323966
GENERAL INFORMATION:
   APPLICANT: Morris, MacDonald
   APPLICANT: Lal, Preeti
   APPLICANT: Diep, Dinh
   TITLE OF INVENTION: METHOD FOR THE IDENTIFICATION OF SEQUENCE POLYMORPHISMS USING
   TITLE OF INVENTION: POLYNUCLEOTIDE SEQUENCE DATABASES, AND SINGLE NUCLEOTIDE
   TITLE OF INVENTION: POLYMORPHISMS IDENTIFIED THEREBY
   FILE REFERENCE: GX-0020-1 P
   CURRENT APPLICATION NUMBER: US/60/323,966
   CURRENT FILING DATE: 2001-09-21
   NUMBER OF SEQ ID NOS: 3246
   SOFTWARE: PERL Program
   SEQ ID NO 1230
   LENGTH: 649
   TYPE: DNA
   ORGANISM: Homo sapiens
   FEATURE:
   NAME/KEY: misc_feature
   OTHER INFORMATION: Incyte ID No: 423998.2
US-60-323-966-1230

Query Match        31.4%;    Score 473.4;   DB 76;   Length 649;
Best Local Similarity   98.6%;   Pred. No. 1.4e-127;
Matches   488;   Conservative    0;   Mismatches   6;   Indels   1;   Gaps   1;
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 05:59:18 ; Search time 1214 Seconds
          (without alignments)
          8884.432 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1 atggctcaagcggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0
Searched: 8166257 seqs, 3573790582 residues

Total number of hits satisfying chosen parameters: 16332514

Post-processing: Minimum Match 0%
                 Maximum Match 100%
Minimum DB seq length: 0
Maximum DB seq length: 200000000
                 Listing first 45 summaries

Database :   Pending_Patents_NA_New:*
     1: /cgn2_6/ptodata/2/pna/PCT_NEW_COMB.seq:*
     2: /cgn2_6/ptodata/2/pna/US06_NEW_COMB.seq:*
     3: /cgn2_6/ptodata/2/pna/US07_NEW_COMB.seq:*
     4: /cgn2_6/ptodata/2/pna/US08_NEW_COMB.seq:*
     5: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq:*
     6: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq2:*
     7: /cgn2_6/ptodata/2/pna/US09_NEW_COMB.seq3:*
     8: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq:*
     9: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq2:*
    10: /cgn2_6/ptodata/2/pna/US10_NEW_COMB.seq3:*
    11: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq:*
    12: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq2:*
    13: /cgn2_6/ptodata/2/pna/US60_NEW_COMB.seq3:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	% Query Match	Length	DB	ID	Description
1	1373.4	91.0	2101	8	US-10-295-027-200	Sequence 200, App
2	1373.4	91.0	2101	10	US-10-094-749-1313	Sequence 1313, Appl
3	1310.2	86.8	1422	10	US-10-275-107-34	Sequence 34, Appl
4	1262.4	83.7	1320	10	US-10-170-235-26112	Sequence 26112, A
5	1260.4	83.5	1320	12	US-60-452-680-2712	Sequence 2712, Ap
6	1259.4	83.5	1395	1	PCT-US02-32850-62	Sequence 62, Appl
7	1259.4	83.5	1395	1	PCT-US02-32850A-62	Sequence 62, Appl
8	596.8	39.5	909	9	US-10-144-771-413	Sequence 413, App
9	496.4	32.9	500	9	US-10-218-140-217	Sequence 217, App
10 c	359.4	23.8	2432	9	US-10-219-382-2114	Sequence 2114, App
11	245	16.2	362	10	US-10-219-382A-2114	Sequence 2114, Ap
12	245	16.2	362	10	US-10-219-382-2114	Sequence 2114, Ap
13	236.4	15.7	601	7	US-09-947-1617	Sequence 1617, App
14	200.6	13.3	201	12	US-60-452-680-42976	Sequence 42976, A
15	200.6	13.3	201	12	US-60-452-680-42977	Sequence 42977, A
16	200.6	13.3	201	12	US-60-452-680-42980	Sequence 42980, A
17	200.6	13.3	201	12	US-60-452-680-42981	Sequence 42981, A
18	200.6	13.3	201	12	US-60-452-680-42982	Sequence 42982, A
19	200.6	13.3	201	12	US-60-452-680-42983	Sequence 42983, A
20	200.6	13.3	201	12	US-60-452-680-42984	Sequence 42984, A
21	199	13.2	201	12	US-60-452-680-42978	Sequence 42978, A
22	188.6	12.5	201	12	US-60-452-680-42979	Sequence 42979, A
23	172.2	11.4	339	10	US-10-219-382-99	Sequence 99, Appl
24	172.2	11.4	339	10	US-10-219-382A-99	Sequence 99, Appl
25	137	9.1	198	6	US-09-912-293-239606	Sequence 239606,
26 c	74	4.9	280	6	US-09-912-293-217609	Sequence 217609,
27	60.6	4.0	1320	10	US-10-156-761-1888	Sequence 1888, Ap
28 c	60.6	4.0	1320	10	US-10-156-761-1	Sequence 1, Appl
29	49.2	3.3	1140	10	US-10-170-235-6358	Sequence 6358, App
30	49.2	3.3	1344	1	PCT-US03-02500-109	Sequence 109, App
31	49.2	3.3	1403	10	US-10-219-051B-5650	Sequence 5650, App
32	49.2	3.3	1415	1	PCT-US02-09576A-114	Sequence 114, App
33	49.2	3.3	1415	9	US-10-342-887-515	Sequence 515, App
34	49.2	3.3	1471	1	PCT-US03-02500-108	Sequence 108, App
35	49.2	3.3	1664	1	PCT-US03-949-016-5120	Sequence 5120, App
36	49.2	3.3	1678	10	US-10-170-235-6468	Sequence 6468, Ap
37	48.2	3.2	1403	10	US-10-425-114-25000	Sequence 25000, A
38	48.2	3.2	1569	10	US-10-425-114-1448	Sequence 1448, Ap
39	46.2	3.1	381	1	PCT-US03-02500-110	Sequence 110, App
40	46	3.0	1429	10	US-09-864-408A-419	Sequence 419, App
41	46	3.0	1576	10	US-09-864-408A-419	Sequence 6359, Ap
42	45.6	3.0	1935	10	US-10-264-213-97	Sequence 97, Appl
43	44	2.9	1323	9	US-10-156-761-6612	Sequence 6612, Ap
44	44	2.9	9025608	10	US-10-156-761-1	Sequence 1, Appl1
45	43.4	2.9	1616	10	US-10-425-114-3715	Sequence 3715, Ap

ALIGNMENTS

RESULT 1
US-10-295-027-200
Sequence 200, Application US/10295027
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynne, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-0125000S
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 200

```
; LENGTH: 2101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-295-027-200

Query Match     91.0%;  Score 1373.4;  DB 8;  Length 2101;
Best Local Similarity   95.8%;  Pred. No. 0;  Mismatches 0;
Matches 1446;  Conservative 0;  Indels 62;  Gaps 1;

Qy          1 ATGGCTCAGCGGTGCGTTTGGCGGTTGGCTGGCCCTTAGTTTTTCCCT      60
              |||||||||||||||||||||||||||||||||||||||||||||||
Db          1 ATGGCTCAGCGGTGCGTTTGGCGTTGGCTGGCCCTTAGTTTTTCCCT      60

Qy         61 ACCGGTCTCCAGATCGANTCGGGCGGCCCGAGGAGGCGTCGGCGAATCCCT     120
              ||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 ACCGGTCTCCAGATCGANTGGGCGGCCCGAGGAGGCGGGGAGCATCAAAGGCGTCGGCGAATCCCT     120

Qy        121 TCTCAGTTCACGCAAAGACAGGAACGCGTCGCGATGAAAAGGTGCCATCCAG     180
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        121 TCTCAGTTCACGCAAAGAGGAACGCGTCGCGATGAAAAGGTGCCATCCAG     180

Qy        181 ANTCCAACAGTGACTTTTTACTCTCAGAAGTCCCAATACTACACAGCCCCTGGCTGAGTTCCGGA     240
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        181 ANTCCAACAGTGACTTTTTAGCTCTCGAGAAGTCCAATACTACACAGCCCCTGAGTTCGGA     240

Qy        241 AAATACATTCANAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGANAGTC     300
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        241 AAATACATTCANAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAAGTC     300

Qy        301 GTGGAAGAGTATAGCCACCTGTTCACTATATCCAAGCTCGGACCCCAGCTTGCAGCCCTAC     360
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCTCGGACCCCAGCTTGCAGCCCTAC     360

Qy        361 CTGCTGANTGGGCTCACTTTGANTGGTGGTGGTCGCCTGCCCCTGAGAAGAGGGTGCCCCCCA     420
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        361 CTGCTGANTGGGCTCACTTTGANTGGTGGTGGTCGCCTGCCCCTGAGAAGAGGGTGCCCCCCA     420

Qy        421 TTCTCTCTGGGTTGGAGCCGTGATGGCGTCATCTATGGTTGGGCACACTGGACGACAAGAAC     480
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TTCTCTGGGTTGGAGCCGTGATGGCGTCATCTATGGTTGGGCACACTGGACGACAAGAAC     480

Qy        481 TCTGTGATGGAGCAATACTGCAGGCCCTTGGAGCTCCTGCTCTGANTCAGGAAAGTACATCCCCCGA     540
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TCTGTGANTGGGCAATACTGCAGGCCTTGGAGCTCCTGCTCTGANTCAGGAAAGTACATCCCCCGA     540

Qy        541 AGATCTTTCTTCATTTCTTCTCTGGGCCACATGATGAGGACAGGGGGCTCAGAGG     600
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGACAGGGGGCTCAGAGG     600

Qy        601 ACTCAGCCCGCTGCTACAGTCAAAGGGGCGTGCTAGCCTCTTCATTGTGACGAGGGGGGC     660
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        601 ATCTCAGCCCCTGCTACAGTCAAAGGGGCGTGCTAGCCTTCATCAGGGACAGGGGGCTCAGGGG     660

Qy        661 TTCATCTTGGANTGANTTTCATTCCTAAACTCAAGAAGCCCATCGCGCTTGANTTGCAGTCTCA     720
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        661 ATCTCAGCCCTGCTACAGTCAAGGGGCCTGAGCCTTCATCAGGGACAGGGGGGC     720

Qy        721 GAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACCCACTCCACTCTTCAGCT     780
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        721 GAGAAGGGTTCATGAACCTCATGCTGCAAGTAAACATGACCCACTCTTCAGCT     780

Qy        781 CCTCCAAAGGAGACAAGCAACCATGGCATCCTTTGCCAGCTTGGAGCAGACA     840
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        781 CCTCCAAAGGAGACAAGCAATCCTTGCCTCGTCTGCGTGCGATTGGAGCAGACA     840

Qy        841 CCAATGCCTATCATATTTGGAAGCAGGGGACAGTGGTGTGACTGTGATTCCAGCAACTGGCAAAT     900
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CCAATGCCTATCATATTGGAAGCGGGGACAGTGGTGGTGTACTGTGATTCCAGCAACTGGCAAAT     900

Qy        901 GAGTTCCTTCCCTCCCTGTCAANTATAATATCCTGAGCAACCACGTGGCTATTTGAAACCACTTATA     960
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        901 CCAATGCCTATACATATTGGAAGCGGGACAGTGGTGGTGTACTGTGATTCCAGCAACTGGCAAAT     960

Qy        961 G-----------------------------------     961
              |
Db        961 GAGTTCCTTCCCTCCCTGTCAANTATAATATCCTGAGCAACCACGTGGCTATTTGAAACCACTTATA     960
```

```
Qy        961 AGCCAGGTTTATGGAGGAGAAAANTCCCTTAACCAATGCAAATAATCAGACACCAGGCACTC     1020
              ||||||||||||||||||||||||||||||||||||||||||||||
Db        962 ---AGGTTTATGGAGGAGAAATCCCTTAACCAATGCAAATAATCAGACACCAGGCACTC     1018

Qy       1021 ACCATANTTCAAAGCACAGGGGTCAAGTTCANTGTCATCCCCCAGTGTGGGCCCAGGCCACAGTC     1080
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1019 ACCATANTTCAAAGCACAGGGGTCAAGTTCANTGTCATCCCCCAGTGTGGGCCCAGGCCACAGTC     1078

Qy       1081 AACTTCCGGATTCACCCTGGACAGACAGTCCAAGTTCANTGTCATCCCCAGTGTGGGCCCACACATC     1140
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1079 AACTTCCGGATTCACCCTGGACAGACAGTCCAAGTTCANTGTCATCCCCAGTGTGGGCCCACACATT     1138

Qy       1141 GTGGCTGATAACAGCAGTCCAGTTCANTGTCATGTGTGTTGAGTGGTCCTTGANTGGCCCTCAGC     1200
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1139 GTGGCTGATAACAGCAGTCCAGTTCANTGTCATGTGTTGAGTGGTCCTTGANTGGCCCTCAGC     1198

Qy       1201 CCTTCTGANTGACAAGGCCTTGGGCTTGGGCTACCAGCTGCTCCGGCGTACAGTCCGCTTTC     1260
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1199 CCTTCTGANTGACAAGGCCTTGGGCTTGGGCTACCAGCTGCTCCGGCGTACAGTCCGCTTTC     1258

Qy       1261 CCGGAAGTCANTATATTACTTGCCCCAGTTACTTCTATTGGCAACACAGACAGCCGANTTCTTT     1320
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1259 CCGGAAGTCANTANTATTACTGCCCCCAGTTACTTCTANTGGCAACACAGACAGCCGANTTCTTT     1318

Qy       1321 ACAAAACCTCACCACTGGCATCTACAGGTTCTCACCCATCTACANTACACACCCTGAAGACTTC     1380
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1319 ACAAAACCTCACCACTGGCATCTACAGGTTCTCACCCATCTACANTACACACCCTGAAGACTTC     1378

Qy       1381 AAACGCATCCANTGGAGTCAACGAGAAAAATCTCAGTTCGACACAGACCAGTTTTCACCTCGCAC     1440
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1379 AAACGCANTCCANTGGAGTCAACGAGAAAAATCTCAGTTCGACACAGACCAAGTGAA     1438

Qy       1441 TTCATCTTTGTGAGTTGANTTCAGANTGCTGACAGACCAGTTTTCTCACCTGCAC     1500
              ||||||||||||||||||||||||||||||||||||||||||||||
Db       1439 TTCATCTTTGAGTTGANTTCAGANTGCTGACAGACCAGTTTTCTCACCTGCAC     1498

Qy       1501 AAACTGTGA     1509
              |||||||||
Db       1499 AAACTGTGA     1507
```

```
RESULT 2
US-10-094-749-1313
; Sequence 1313, Application US/10094749
; GENERAL INFORMATION:
; APPLICANT: ISOGAI, TAKAO
; APPLICANT: SUGIYAMA, TOMOYASU
; APPLICANT: OTSUKI, TETSUJI
; APPLICANT: WAKAMATSU, AI
; APPLICANT: SATO, HIROYUKI
; APPLICANT: ISHII, SHIZUKO
; APPLICANT: YAMAMOTO, JUN-ICHI
; APPLICANT: ISONO, YUUKO
; APPLICANT: HIO, YURI
; APPLICANT: OTSUKA, KAORU
; APPLICANT: NAGAI, KEIICHI
; APPLICANT: IRIE, RYOTARO
; APPLICANT: TAMECHIKA, ICHIRO
; APPLICANT: SEKI, NAOHIKO
; APPLICANT: YOSHIKAWA, TSUTOMU
; APPLICANT: OTSUKA, MOTOYUKI
; APPLICANT: NAGAHARI, KENJI
; APPLICANT: MASUHO, YASUHIKO
; TITLE OF INVENTION: NOVEL FULL-LENGTH cDNA
; FILE REFERENCE: 084335/0160
; CURRENT APPLICATION NUMBER: US/10/094,749
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 60/350,435
; PRIOR FILING DATE: 2002-01-24
; PRIOR APPLICATION NUMBER: JP 2001-328381
; PRIOR FILING DATE: 2001-09-14
; NUMBER OF SEQ ID NOS: 3381
; SOFTWARE: Patentin Ver. 2.1
```

; SEQ ID NO 1313
; LENGTH: 2101
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-094-749-1313

Query Match   91.0%; Score 1373.4; DB 10; Length 2101;
Best Local Similarity   95.8%; Pred. No.0;
Matches 1446; Conservative 0; Mismatches 1; Indels 62; Gaps 1;

```
Qy      1 ATGGCTCAGCGGGTGCGTTTGCGTGCTGGGCGCTGGTGGGCTGGGCCTGGTGGGCTGGCGTGGGCTGCTCCTAGTTTTCCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ATGGCTCAGCGGGTGCGTTTGCGTGCTGGGCGCTGGTGGGCTGGGCCTGGTGGGCTGGCGTGGGCTGCTCCTAGTTTTCCCT 120

Qy     61 ACCGTCTCCAGATCGATCGATGGGCCCGAGGAGCGCGGGAGCATCAAAAGGCCGTTCGCGGCGAATCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 ACCGTCTCCAGATCGATCGATGGGCCCGAGGAGCGCGGGAGCATCAAAAGGCCGTTCGCGGCGAATCCCT 180

Qy    121 TCTCAGTTCACCAAAGAGAGGAACGCGGTCTCGGGATGAAAGAGGGCGTTCGAAAGGTGTGCCATTCCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TCTCAGTTCACCAAAGAGAGGAACGCGGTCTCGGGATGAAAGAGGGCGTTCGAAAGGTGTGCCATTCCAG 240

Qy    181 ATTCCAACAGTGACTTTTAGCTCTCGAGAAGTCCAATACTACAGCCCCTGGCTGGCTGAGTTCCGGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 ATTCCAACAGTGACTTTTAGCTCTCGAGAAGTCCAATACTACAGCCCCTGGCTGGCTGAGTTCCGGA 300

Qy    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCGACCAGCTTTTATTCAGCATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCGACCAGCTTTTATTCAGCATGAAGTC 360

Qy    301 GTGGAGAGAGTATAGCCACCTGTTCACTACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 GTGGAGAGAGTATAGCCACCTGTTCACTACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 420

Qy    361 CTGCTGATGGCTCACTTTGATGTGGTGCCTCGCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CTGCTGATGGCTCACTTTGATGTGGTGCCTCGCCCCTGAAGAAAGGCTGGGAGGTGCCCCA 480

Qy    421 TTCTCTCGGGTTGGAGCCGTCGATGGCGTCATCTATGGTCGGGCCACACTGGACGACAAGAAC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TTCTCTCGGGTTGGAGCCGTCGATGGCGTCATCTATGGTCGGGCCACACTGGACGACAAGAAC 540

Qy    481 TCTGTGATGGGCATTACTGCCAGGCCTTGGAGCTCCTGATCAGGAGAGTACAATCCCCCGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 TCTGTGATGGGCATTACTGCCAGGCCTTGGAGCTCCTGATCAGGAGAGTACATCCCCCGA 600

Qy    541 AGAGCTTCTTCATTTCTCTGGGGCCCATGATGAGGAGTCATCAGGAGGGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 AGAGCTTCTTCATTTCTCTGGGGCCCATGATGAGGAGTCATCAGGAGGGGGCTCAGAGG 660

Qy    601 ATCTCAGCCCTGCTACAGTCAAGGGGGTCCAGCTACCCTTCATTGTTGGACAGGGGGGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 ATCTCAGCCCTGCTACAGTCAAGGGGGTCCAGCTACCCTTCATTGTTGGACAGGGGGGC 720

Qy    661 TTCATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGCCCTTGATTCCAGTCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TTCATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGCCCTTGATTCCAGTCTCA 780

Qy    721 GAGAGAGGGTTCCATGAACCTGCTGCAAGTAAACATGACTTCAGGCCGCACTCTTCAGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 GAGAGAGGGTTCCATGAACCTGCTGCAAGTAAACATGACTTCAGGCCGCACTCTTCAGCT 840

Qy    781 CCTTCCAAAAGGAGACAAGCATTGGGATCCTTGCACTGTGTCACCTGTCTTCGCACCT 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCTTCCAAAAGGAGACAAGCATTGGGATCCTTGCACTGTGTCACCTGTCTTCGCACCT 900

Qy    841 CCAATGCCTATCATATATTGGAAAGCGGGACAGTGGTGTGACTGCTGTATTGCAGCAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCAATGCCTATCATATATTGGAAAGCGGGACAGTGGTGTGACTGCTGTATTGCAGCAAAT 960

Qy    901 GAGTTTCCCTTCCCTGTCAATAATAATCCTGAGCAACCCATGGCTATTTGAACCACTATA 960
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||------
Db    961 G------- 961
```

RESULT 3
US-10-275-107-34
; Sequence 34, Application US/10275107
; GENERAL INFORMATION:
; APPLICANT: PLOWMAN, GREGORY D.
; APPLICANT: WHYTE, DAVID
; APPLICANT: SUDARSANAM, SUCHA
; APPLICANT: MANNING, GERARD
; APPLICANT: CAENEPEEL, SEAN R.
; APPLICANT: PAYNE, VILLA
; TITLE OF INVENTION: NOVEL PROTEASES
; FILE REFERENCE: 038602/1479
; CURRENT APPLICATION NUMBER: US/10/275,107
; CURRENT FILING DATE: 2003-11-03
; PRIOR APPLICATION NUMBER: PCT/US01/14431
; PRIOR FILING DATE: 2001-05-04
; PRIOR APPLICATION NUMBER: 60/201,879
; PRIOR FILING DATE: 2000-05-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: Patentin Ver. 2.1
;
; SEQ ID NO 34
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-275-107-34

Query Match   86.8%; Score 1310.2; DB 10; Length 1422;
Best Local Similarity   94.0%; Pred. No.0;
Matches 1419; Conservative 0; Mismatches 3; Indels 87; Gaps 2;

```
Qy    961 AGCAGGTTTATGGAGAGAGAAATCCCTTAACCAATGCAATAATCAGGAGACCACCAGGCACTC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    962 ---AGGTTTATGGAGAGAGAAATCCCTTAACCAATGCAATAATCAGGAGACCACCAGGCACTC 1018

Qy   1021 ACCATAATTCAAAGCAGGGGGTCAGGTCGTCGGGGTTGGGGGCCAGGGCCCAGGGCCACAGTC 1080
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1019 ACCATAATTCAAAGCAGGGGGTCAGGTCGTCGGGGTTGGGGGCCAGGGCCCAGGGCCACAGTC 1078

Qy   1081 AACTTCCGGATTCACCCCTGGACAGACAGACAGTCCAAGGAGGTCCTAGAACTCACGAAGAACATT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1079 AACTTCCGGATTCACCCCTGGACAGACAGACAGTCCAAGGAGGTCCTAGAACTCACGAAGAACATT 1138

Qy   1141 GTGGGCTGATAACAGAGTCCAGTTCCACATGTGTTGGAGGCGTTTGACCCCCTCCCCGTCAGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1139 GTGGGCTGATAACAGAGTCCAGTTCCACATGTGTTGGAGGCGTTTGACCCCCTCCCCGTCAGC 1198

Qy   1201 CCTTCTGATGACAAGGCCCTTGGGCTTACCAGCTGCTCCGCGAGACGTACCACGTCGTCTTC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1199 CCTTCTGATGACAAGGCCCTTGGGCTTACCAGCTGCTCCGCGAGACGTACCACGTCGTCTTC 1258

Qy   1261 CCGGAAGTCAAATATTACTGCCCCAGTTACCTTTCTTATTGGGCAACACAGACACAGCCGGATTCTTT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1259 CCGGAAGTCAAATATTACTGCCCCAGTTACCTTTCTTATTGGGCAACACAGACACAGCCGGATTCTTT 1318

Qy   1321 ACAAACCTCACCACTGGCATCTACAGGTTCTACCCCCATCACATACACGGTGAAGACTTC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1319 ACAAACCTCACCACTGGCATCTACAGGTTCTACCCCCATCACATACACGGTGAAGACTTC 1378

Qy   1381 AAACGCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTATGAGAGACCCAAGTGAAA 1440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1379 AAACGCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTATGAGAGACCCAAGTGAAA 1438

Qy   1441 TTCATCATTTGGAGTTGATTCAGAATGCTCGACACAGACCAGGAGGCCAGTTTTCTCACCTGCAC 1498
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1439 TTCATCATTTGGAGTTGATTCAGAATGCTCGACACAGACCAGGAGGCCAGTTTTCTCACCTGCAC 1498

Qy   1501 AAACTGTGA 1509
          |||||||||
Db   1499 AAACTGTGA 1507
```

SEQ ID NO 34
; LENGTH: 1422
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-275-107-34

Query Match   86.8%; Score 1310.2; DB 10; Length 1422;
Best Local Similarity   94.0%; Pred. No.0;
Matches 1419; Conservative 0; Mismatches 3; Indels 87; Gaps 2;

```
Qy      1 ATGGCTCAGCGGGTGCGGTTGCGTTGCGGTGCTGGGCGCTCCTCCTAGTTTTCCCT 60
```

Tue Jul 1 11:49:15 2003   us-10-014-896-1.rnpn   Page 4

```
RESULT 4
US-10-170-235-26112
Sequence 26112, Application US/10170235

GENERAL INFORMATION:
  APPLICANT: VENTER, J. Craig
  TITLE OF INVENTION: KITS, SUCH AS NUCLEIC ACID ARRAYS, COMPRISING A MAJORITY OF HU
  TITLE OF INVENTION: TRANSCRIPTS, FOR DETECTING EXPRESSION AND OTHER USES THEREOF
  FILE REFERENCE: CL001380
  CURRENT APPLICATION NUMBER: US/10/170,235
  CURRENT FILING DATE: 2003-03-17
  NUMBER OF SEQ ID NOS: 42514
SEQ ID NO 26112
  LENGTH: 1320
  TYPE: DNA
  ORGANISM: HUMAN
US-10-170-235-26112

Query Match           83.7%;  Score 1262.4;  DB 10;  Length 1320;
Best Local Similarity  99.8%;  Pred. No. 0;
Matches 1274; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
```

(Nucleotide sequence alignment data follows, Qy vs Db rows, positions 1–1422)

; APPLICANT: GRUPE, Andrew
; TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
; TITLE OF INVENTION: ALZHEIMER'S DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001450
; CURRENT APPLICATION NUMBER: US/60/452,680
; CURRENT FILING DATE: 2003-03-07
; NUMBER OF SEQ ID NOS: 116213
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2712
; LENGTH: 1320
; TYPE: DNA
; ORGANISM: Homo sapiens
US-60-452-680-2712

Query Match   83.5%; Score 1260.4; DB 12; Length 1320;
Best Local Similarity   99.2%; Pred. No. 0; Indels 1; Gaps 1;
Matches 1266; Conservative 9; Mismatches 0;

RESULT 5
US-60-452-680-2712
; Sequence 2712, Application US/60452680
; GENERAL INFORMATION:
; APPLICANT: CARGILL, Michele

```
RESULT 6
PCT-US02-32850-62
Sequence 62, Application PC/TUS0232850
GENERAL INFORMATION:
   APPLICANT: INCYTE GENOMICS, INC.
   APPLICANT: RAMKUMAR, Jayalaxmi
   APPLICANT: GORVAD, Ann E.
   APPLICANT: BAUGHN, Mariah R.
   APPLICANT: EMERLING, Brooke M.
   APPLICANT: YANG, Junming
   APPLICANT: LEE, Soo Yeun
   APPLICANT: TRAN, Uyen K.
   APPLICANT: BECHA, Shanya D.
   APPLICANT: DUGGAN, Brendan M.
   APPLICANT: LEE, Ernestine A.
   APPLICANT: GRIFFIN, Jennifer A.
   APPLICANT: LI, Joana X.
   APPLICANT: SPRAGUE, William W.
   APPLICANT: HAFALIA, April J.A.
   APPLICANT: CHAWLA, Narinder K.
   APPLICANT: LEHR-MASON, Patricia M.
   APPLICANT: KABLE, Amy E.
   APPLICANT: YUE, Henry
   APPLICANT: MARQUIS Joseph P.
   APPLICANT: YAO, Monique G.
   APPLICANT: RICHARDSON, Thomas W.
   APPLICANT: TANG, Y. Tom
   APPLICANT: JIN, Pei
   APPLICANT: CHIEN, David
   APPLICANT: BHATIA, Umesh G.
   APPLICANT: BURRILL, John D.
   APPLICANT: LEE, Sally
   APPLICANT: BLAKE, Julie J.
   APPLICANT: HO, Anne
   APPLICANT: ZHENG, Wenjin
   TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE PROTEINS
   FILE REFERENCE: PF-1237 PCT
   CURRENT APPLICATION NUMBER: PCT/US02/32850
   CURRENT FILING DATE: 2002-10-11
```

```
   PRIOR APPLICATION NUMBER: US 60/329,689
   PRIOR FILING DATE: 2001-10-12
   PRIOR APPLICATION NUMBER: US 60/335,703
   PRIOR FILING DATE: 2001-10-25
   PRIOR APPLICATION NUMBER: US 60/348,887
   PRIOR FILING DATE: 2001-11-09
   PRIOR APPLICATION NUMBER: US 60/334,145
   PRIOR FILING DATE: 2001-11-28
   PRIOR APPLICATION NUMBER: US 60/337,451
   PRIOR FILING DATE: 2001-12-06
   PRIOR APPLICATION NUMBER: US 60/340,584
   PRIOR FILING DATE: 2001-12-14
   NUMBER OF SEQ ID NOS: 80
   SOFTWARE: PERL Program
SEQ ID NO 62
   LENGTH: 1395
   TYPE: DNA
   ORGANISM: Homo sapiens
   FEATURE:
   NAME/KEY: misc_feature
   OTHER INFORMATION: Incyte ID No: 3118830CB1
PCT-US02-32850-62

Query Match        83.5%;   Score 1259.4;   DB 1;  Length 1395;
Best Local Similarity  95.5%;   Pred. No. 0;  Matches 1332;  Conservative 0;  Mismatches 1;  Indels 62;  Gaps 1;
```

```
Qy         1 ATGGCTCAGCGTGCGGTTGCGTTTGCGTGTGCTGGTGCGCCCTTTGTCGGCCTATTTCCTTTTTCCTT   60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        63 ATGGCTCAGCGGTGCGTTTGCGTTTGCCGTTGCCCTGGGTCGGCTATGCTGTTTTTTCCT  122

Qy        61 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGGAGCATCAAAGGCGGCGGGAAATCCCT  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       123 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGGAGCATCAAAGGCGGCGGGAATCCCT  182

Qy       121 TCTCAGTTCAGCAAAGAAGGAACGCGTTCGCCGGATGAAAGAGGGCGCTGAAAGGTGCCATCCAG  180
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       183 TCTCAGTTCAGCAAAGAAGGAACGCGTTCGCCGGATGAAAGAGGGCGCTGAAAGGTGCCATCCAG  242

Qy       181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGCTGAGTTCGGA  240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       243 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGCTGAGTTCGGA  302

Qy       241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCAGCTTTATCCAGCATGAAGTC  300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       303 AAATACATTCATAAAAGTCTTTCCTACAGTGGTCAGCAGCTTTATCCAGCATGAAGTC  362

Qy       301 GTGGAAGAGTATAGCCAACCTGTTCACTATCCAAGCGCTCGGACCCCAGCTTGCAGCCTAC  360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       363 GTGGAAGAGTATAGCCAACCTGTTCACTATCCAAGCGCTCGGACCCCAGCTTGCAGCCTAC  422

Qy       361 CTGCTGATGGCTCACTTTGATGTGGTTGGCCTGCCCCCTGAGAGAAGGCTGGGAGGTGCCCCCA  420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       423 CTGCTGATGGCTCACTTTGATGTGGTTGGCCTGCCCCCTGAGAGAAGGCTGGGAGGTGCCCCCA  482

Qy       421 TTCTCATGGGGTTGGGAGGTGATGGCGTCATCGAGGCAGACAAGAAC  480
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       483 TTCTCATGGGGTTGGGAGGTGATGGCGTCATCTATGGTTCGGGCACACTGACGACAAGAAC  542

Qy       481 TCTGTGGAATGGCATTACTGCAGGCCCTTGGAGCTCCTGGATCAGGAAGTACATCCCCCGA  540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       543 TCTGTGGAATGGCATTACTGCAGGCCCTTGGAGCTCCTGGATCAGGAAGTACATCCCCCGA  602

Qy       541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGACAGAGGGGCTCAGAGG  600
             |||||||||||||||||||||||||||||||||||||||||||||||||
Db       603 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGACAGAGGGGCTCAGAGG  662

Qy       601 ATCTCAGCCCTGCTACAGTCAAAGGGGCCGTCCAGCTAGCCTTCATTGTGGACGAGGGGC  660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       663 ATCTCAGCCCTGCTACAGTCAAAGGGGCCGTCCAGCTAGCCTTCATTGTGGACGAGGGGC  722

Qy       661 TTCATCTTGGATGATTTCATTTCCTAACTTCAAGAAGCCCATCGCGCCTTGATTGCAGTCTCA  720
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Qy       840 ACCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAA  899
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       885 ACCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACWGTATTGCAGCAACTGGCAAA  944

Qy       900 TGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAAACCCATGGGCTATTTGAACCACTTAT  959
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       945 TGAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAAACCCATGGCTATTTGAACCACTTAT  1004

Qy       960 AAGCAGGTTTATGGAGGAGAAATCCCTTTAACCAATGCAATAATCAGGACCACCAGGCACT  1019
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1005 AAGCAGGTTTATGGAGGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCAGGCACT  1064

Qy      1020 CACCAATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCCCACAGT  1079
             ||||||:||||||||||||||||||||:||||||||||||||||||||||||||||
Db      1065 CACCATATTCAAAGCARGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCCCACACT  1124

Qy      1080 CAACTTCCGGATTCACCCTGGACAGACACAGTCCCAGAGGTTCCTAGAACTCACGAAGAACAT  1139
             |||||||||||||||||||||||||||||||||||||||:||||:::
Db      1125 CAACTTCCGGATTCACCCTGGACAGACAGTTCCAAGAGGTTCCTAGAACTCACGAARAACAY  1184

Qy      1140 TGTGGCTGATAAACAGAGTCCAGTTCCATGTGTTGAGTGTGGTGAGTGCCTTTGACCCCCTCCCCGTCAG  1199
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1185 TGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGACCCCCTTTGACCCCCTCCCCGTCAG  1244

Qy      1200 CCCTTCTGATGACAAGGCCTTGGGGCTACCAGCTGCCCCAGCAACGGTACAGTCCGTCTCTT  1259
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1245 CCCTTCTGATGACAAGGCCTTGGGCTACCAGCTGCTCCCAGCAACGGTACAGTCGGTCTCTT  1304

Qy      1260 CCCGGAAGTCAATAATT  1275
             |||||||||||||||||
Db      1305 CCCGGAAGTCAATAATT  1320
```

RESULT 7

PCT-US02-32850A-62

Sequence 62, Application PC/TUS0232850A

GENERAL INFORMATION:

APPLICANT: INCYTE GENOMICS, INC.
APPLICANT: RAMKUMAR, Jayalaxmi
APPLICANT: GORVAD, Ann E.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: EMERLING, Brooke M.
APPLICANT: YANG, Junming
APPLICANT: LEE, Soo Yeun
APPLICANT: TRAN, Uyen K.
APPLICANT: BECHA, Shanya D.
APPLICANT: DUGGAN, Brendan M.
APPLICANT: LEE, Ernestine A.
APPLICANT: GRIFFIN, Jennifer A.
APPLICANT: LI, Joana X.
APPLICANT: SPRAGUE, William W.
APPLICANT: HAFALIA, April J.A.
APPLICANT: CHAWLA, Narinder K.
APPLICANT: LEHR-MASON, Patricia M.
APPLICANT: KABLE, Amy E.
APPLICANT: YUE, Henry
APPLICANT: MARQUIS Joseph P.
APPLICANT: YAO, Monique G.
APPLICANT: RICHARDSON, Thomas W.
APPLICANT: TANG, Y. Tom
APPLICANT: JIN, Pei
APPLICANT: CHIEN, David
APPLICANT: BHATIA, Umesh G.
APPLICANT: BURRILL, John D.
APPLICANT: LEE, Sally
APPLICANT: BLAKE, Julie J.
APPLICANT: HO, Anne
APPLICANT: ZHENG, Wenjin

TITLE OF INVENTION: PROTEIN MODIFICATION AND MAINTENANCE MOLECULES

SEQ ID NO 62
LENGTH: 1395
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No: 3118830CB1

PCT-US02-32850A-62

Query Match 83.5%; Score 1259.4; DB 1; Length 1395;
Best Local Similarity 95.5%;
Matches 1332; Conservative 0; Mismatches 0; Indels 62; Gaps 1;

```
Qy    1 ATGGCTCAGCGGTGCGGTTTGCCGTGCCGATGAAGAAGAGGCCGGCCTGAAAGGTGCCATCCAG   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   63 ATGGCTCAGCGGTGCGGTTTGCCGTGCCGATGAAGAAGAGGCCGGCCTGAAAGGTGCCATCCAG  122

Qy   61 ACCGGTCTCCAGATCGATGGGCCCGGAGGAGGCGGGAGCATCAAAAGGGGGTCGGGAAATCCCT  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  123 ACCGGTCTCCAGATCGATGGGCCCGGAGGAGGCGGGAGCATCAAAAGGGGGTCGGGAAATCCCT  182

Qy  121 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAGAAGAGGCCCTGAAAGGTGCCATCCAG    180
          |||||||||||||||||||||||||||||||||||||||||
Db  183 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAGAAGAGGCCCTGAAAGGTGCCATCCAG     242

Qy  181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGGAGTTCGGA    240
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  243 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA     302

Qy  241 AAATACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC  300
          |||||||||||||||||||||||||||||||||||||||||||
Db  303 AAATACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC   362

Qy  301 GTGGAAGAGTATAGCCACCTGTTCACTATACCAAGGCTCGGACACCCAGCACCCTAC    360
          |||||||||||||||||||||||||||||||||||||||||||
Db  363 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACACCCAGCACCCTAC     422

Qy  361 CTGCTGATGGGCTCACTTTGAATGTGGTGGCTGCCCCTGAAGAAGAGGGCTTGGGAGGTG    420
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  423 CTGCTGATGGGCTCACTTTGATGTGGTGGCTGCCCCTGAAGAAGAGGGCTTGGGAGGTG     482

Qy  421 TTCTCTGGGTTGGACGGTGATGGGCGGTCATCTATGGTCGGGCGCACACTGGACGACAAGAAC  480
```

; TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
; FILE REFERENCE: CL001321
; CURRENT APPLICATION NUMBER: US/10/144,771
; CURRENT FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 413
; LENGTH: 909
; TYPE: DNA
; ORGANISM: HUMAN
US-10-144-771-413

RESULT 8
US-10-144-771-413
; Sequence 413, Application US/10144771
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig

Query Match 39.5%; Score 596.8; DB 9; Length 909;
Best Local Similarity 80.5%; Pred. No. 1.1e-163;
Matches 729; Conservative 0; Mismatches 162; Indels 15; Gaps 2;

```
Qy      1 ATGGCTCAGCGGTGCGTTGCCTGGCCGTTCCTGCTGGCCCTGCTGGCGGTGGCCTGCGCCTCCTAGTTTTCCCT 60
          ||| || ||||| | || || ||| | |  |  |  |||| || || | | ||  | |   || | |||||||||
Db      1 ATGGCTGAGCTACTTGCTAGCTTGCTTGCCCGCCTGGGCCCTGGGCTGCCTGCCTTCTTTTTTTCGT 75

Qy     61 ACGGTCTCCCAGATCGGATCGGGGCGGGCCCGGAGGGAGGGAGGGAGCATCAAAGGGGTCGGGAATCCT 120
          |  | |||   ||| | ||| || ||||||| ||| ||||||| ||| ||| || ||||||| ||||||
Db     16 ATGGCTGAGCTACTTGCTAGCTTGCTTGCCCGCCTGGGCCCTGGGCTGCCTGCCTTCTTTTTTCGT 75

Qy    121 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAGAAAGAGAAGGTGCCATCCAG 180
          ||| |||||||  || ||| || |||||   |||| ||| ||| |||||||||||
Db    136 TCCCAGTTCAGCGAGGAGGACGCGTGCGCTAATAAAGAGAGGGTGCGCTAAAGGTGCCATCCAG 195

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATAACTACACAGCGCTGGTTCGGA 240
          ||||||| ||||| ||||||||||||| ||||||| || |  |||||| ||| ||||
Db    136 ATTCCAACAGTGTCTTTCAGCGAGGAATCCAACACCAGCCCTTGCTTGGATTCGGA 255

Qy    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCAGACTTTATTCAGCAATGAAGTC 300
          |||| ||| |||||| |   |  ||||| ||||| |||| ||  |||||| ||| ||||
Db    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCAGCTTTATTCAGCAATGAAGTC 300

Qy    301 GTGGAACAAGTATAGCCACCGTTCACTACTATCCGAACCCGACCCCCTAC 360
          |||| |||||||||| ||||  |||   |||||||  ||||||||||||
Db    256 GAATAATATCCGCAA------AGTGTTCCACAGCAGCCTTGCTCCAAACATGAAGTC 303

Qy    361 CTGCTGAATGGGCTCACTTTGATGTGGTGGCCTGCCCCCTCCCCCTCCCCCA 420
          ||| | |||||| ||||||||  | |||| | |  |||||||||||
Db    364 ATGCTGATGGGCTCACATTGATGTGGTTGTTCGGGGGACAGAAGAATGGGGTCCCCG 423

Qy    421 TTCTCTGGGTTGGAGCGGTGATTGGGCTCATCTATTGGTCGGGGCACACTGGACGACAAGAAC 480
          ||||||||||||||| | |||| |||||  || || || || || ||||| ||||| || ||
Db    424 TTCTCAGGCCTGGAACCCAATGGCTTCATCTACTCTATGGCGCGGGGGACTGGACGACAAAAAC 483

Qy    481 TCTGTGATGGCATTACTGCAGGCCCTTGGAGCGTCCTGCTGGAGGAAGTACATCCCGA 540
          |||||||| || ||||||||||||| | ||||| | | |||| || |||||||||||
Db    481 TCTGTGATGGCATTACTGCAGGCCCTTGGAGCGTCCTGCTGGAGAAGTACATCCCGA 540

Qy    541 AGATCTTTCTTCATTTCTCTGGGCCATGAATGAGGAGACAGGGGTCTCAGAGG 600
          ||| ||||||| |  | || ||||| ||| |  |||| |||||| ||||||||
Db    484 AGATCTTTCTTCATTTCTCTGGGCCAATGAGAGCTTCTCGTTGTTGGACGAGGG 543

Qy    601 AGATCTTTCTTCATTTCTCTGGGCCCCATGATGAGGAGACAGGGGGTCAGGAG 660
          ||| ||||||| |  | || ||||| || |  |||| |||||| ||||||||
Db    544 AGATCTTCATTCTCGGGGATGAGCAGCTAGGGGTCCAGGAGGGTCGGCCTCAG 603

Qy    661 TTCATCTTGATCGATTCATTTCATTTCTCTAACTTCAAGAAGCCCATCTCCCTTGATTCGCAGGAGTCTCA 720
          || |||||||| |||||||||||||||| ||||||||||| |||||||||||||||||||| | ||||||
Db    604 AAGATCTCAGCACTCTTACAGGCAAGGGGTCACCTTCCTTTGGGGATGAAGGG 663

Qy    721 GAGAAGGGTTCCATGAACCTCATGGCTGCAAGTAAACATGACTTCAGGCCCACTCTTCAGCT 780
          ||| ||| |||||||| || ||| || ||||||||| ||||||||||||||||| ||||
Db    658 GGCTTCATCTTGGATGAATTTCATTCCTAACTTCAAGAAGCCCATCGCCGTTGATTCAGTC 717

Qy    781 CCTCCAAAGGAGCAAGCAATTGGGCCATCCTTGCGCTGCTCGCTGGTAGGCAGCAGACA 840
          ||||||||||||||| |||||||||| |||| | |||||| || || || ||||||||
Db    718 TCAGAGAAGGGTTCCATGAACCTGCAAGTAAACATGACTTCAGGCACTCTTCA 777

Qy    841 CCAATGCCTATCATATTTGGAAGCGGGGACAGTGGTGACTGTATTGCCAAAT 900
          |||||||||||||||||||||||||||| ||||||||| || ||| ||||
Db    724 ACTGAGAAGGGTTGCCCCTTGACCTGCAAGGTAAACAATGGACTCCAGAAGGT 783

Qy    901 GAGTTTCCCTTCCCTGTCAAATATAATCCTGAGCAACCACAGGCTATTTGAACCACTTATA 960
          |||||||||||||||||| ||| || ||||||||
Db    778 GCTCCTCCAAAAGGAGCACAAGCATTGGCCATCTCTGCAGCGTCCTGTGCAGCGGGGGCAG 837

Qy    961 AGCAGGTTTATTGGAGAGAAATCCCCTTAACCAAATGCAATAATCAGGACCCACTC 1020
          ||| || ||  || || ||||||||||||||| |||||||| || |||||||||||
Db    784 GCTCCCCCAAAAGGAGCACAAGCATTCTTTCTTGCCGGCCTGTCAGCCGACTGGAGCAG 843

Qy   1021 ACCATAATTCAAAGCACAGGGGGTCAAGTTCAAATGTCATCCCCCCAGTGCCCAGTC 1080
          ||||| |||||||||||||| || || ||||| |||| ||||||||||||| ||||
Db    838 ACACCAATGCCTATCATATTTGGAAGCGGGGACAGTGGTGACTGTATTGCAGCAACTGGCA 897

Qy   1081 AACTTCCGGATTCACCCTGACAGACAGGTCCTAGAACTCACGAGAGAACATT 1140

Qy   1141 GTGGCTGATAACAGAGTCCAGTTCCATGTGTGTTGAGTGCCTTTGACCCCCTCCCCGTCAGC 1200

Qy   1201 CCTTCTGATGACAGTCACCTGACGACTGTCTCCGCCGACCGTACAGTCCGATTCTTT 1260

Qy   1261 CCGGAAGTCAATATATTACTGCCCAGTTGCTTACTTGCTTCTATTGGCAACACAGCACCGATT 1320

Qy   1321 ACAAACCTCACCACTCGGCATCTACAGGTTCTACCCCATCTACACAGCCTGAAGACTTC 1380

Qy   1381 AAACGGCATCCATGGA 1395
          ||||||||||||||||
Db   1381 AAACGGCATCCATGGA 1395
```

```
Db    844 ACACCAAATGCCGAATATGTTTTGGAGGAGGAGGCCAATTGAGGAAGACAATGAAGCTACTGGCA 903

Qy    898 AATGAG 903
          ||||||
Db    904 AATGAG 909
```

RESULT 9
US-10-218-140-217
; Sequence 217, Application US/10218140
; GENERAL INFORMATION:
;   APPLICANT: Leach, Martin D.
;   APPLICANT: Shinkets, Richard A.
;   TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES AND POLYPEPTIDES
;   TITLE OF INVENTION: ENCODED THEREBY
;   FILE REFERENCE: 15966-543 CON
;   CURRENT APPLICATION NUMBER: US/10,218,140
;   CURRENT FILING DATE: 2002-08-12
;   PRIOR APPLICATION NUMBER: 09/540,763
;   PRIOR FILING DATE: 2000-03-30
;   PRIOR APPLICATION NUMBER: 60/127,728
;   PRIOR FILING DATE: 1999-04-05
;   PRIOR APPLICATION NUMBER: 60/127,636
;   PRIOR FILING DATE: 1999-04-02
;   PRIOR APPLICATION NUMBER: 60/127,607
;   PRIOR FILING DATE: 1999-03-31
;   NUMBER OF SEQ ID NOS: 6322
;   SOFTWARE: CuraTator Version 1.0
; SEQ ID NO 217
;   LENGTH: 500
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(499)
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(1)
;   OTHER INFORMATION: "n" = "a", "c", "t" or "g"
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (2)..(2)
;   OTHER INFORMATION: "n" = "a", "c", "t" or "g"
US-10-218-140-217

Query Match   32.9%;   Score 496.4; DB 9; Length 500;
Best Local Similarity   99.8%;   Pred. No. 2.1e-134;
Matches 497; Conservative   0; Mismatches 1; Indels   0; Gaps   0;

```
Qy    141 ACGCGTCGCGATGAAAGAGGCCGCTGAAAGGTGCCATCCAGAGTTCCAACAGTGACTTTTAG 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 ACGCGTCGCGATGAAAGAGGCCGCTGAAAGGTGCCATCCAGAGTTCCAACAGTGACTTTTAG 62

Qy    201 CTCTGAGAAGTCCAATACTACAGAGCCCTGGCTGCGTTCGGGAAAAATACATTCATAAAGTCTT 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     63 CTCTGAGAAGTCCAATACTACAGAGCCCTGGCTGCGTTCGGGAAAAATACATTCATAAAGTCTT 122

Qy    261 TCCTACAGTGGTCAGCACCAGCAGCTTTATCCAGCATGAGAGTCGTGGAAGAGTATAGCCACCT 320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TCCTACAGTGGTCAGCACCAGCAGCTTTATCCAGCATGAGAGTCGTGGAAGAGTATAGCCACCT 182

Qy    321 GTTCACTATCCAAGGCCTCGGACCCCAGCTTGCAGCCTCACTGCTCACTTTGA 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCTCACTGCTCACTTTGA 242

Qy    381 TCTGGTGCTCCTGCCCCTCGAAGAAGGGCTGGGAGTGGGCATGCCCCCTCCCCATTCTCTCTTGGGGTTGGAGCCTGA 440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TGTGGTGCCTGCCCCTGAAGAAGGGCTGGGAGTGGGCCCCCCATTCTCTCGGGTTGGAGCCTGA 302

Qy    441 TGGCGTCTCATCTATAGGTCGGGGGACACTGGACGACGACGACAAGAACTCTGTGATGGCATTACTGCA 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TGGCGTCATCTATAGGTTGGGGGACACTGGACGACAAGAAACTCTGTGATGGCATTACTGCA 362
```

RESULT 10
US-10-144-771-412
; Sequence 412, Application US/10144771
; GENERAL INFORMATION:
;   APPLICANT: VENTER, J. Craig
;   TITLE OF INVENTION: HUMAN GENOME DISCOVERY SYSTEM AND USES THEREOF
;   FILE REFERENCE: CL001321
;   CURRENT APPLICATION NUMBER: US/10,144,771
;   CURRENT FILING DATE: 2002-05-15
;   NUMBER OF SEQ ID NOS: 47235
; SEQ ID NO 412
;   LENGTH: 2432
;   TYPE: DNA
;   ORGANISM: HUMAN
US-10-144-771-412

Query Match   23.8%;   Score 359.4; DB 9; Length 2432;
Best Local Similarity   77.0%;   Pred. No. 5.8e-94;
Matches 438; Conservative   0; Mismatches 131; Indels   0; Gaps   0;

```
Qy    941 GGCTATTTGAACCACTATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAATAA 1000
          ||||||||| ||| || |||||||||||| || | |||||||| |||| ||||||||
Db      1 GGCTATTTCATCCCATTGTGAGCAGGATAATAGTGGAGAACCCCCATAACAAATGCGGCTGG 60

Qy   1001 TCAGGACGACCACGGGGCCACTCACCATATTCAAAGCAGGGGTCAAGTTCAAGTCATCCCCC 1060
          ||| |||||| ||| |||||| ||| ||| ||| || || ||| || || ||||||||
Db     61 TCCGAACTACCACCAGCCCTCACCATGTCAATGCAGGAATCAAATGAAAGTCATCCCTC 120

Qy   1061 CAGTGCCCCAGGCCACAGTCAAACTTCCGGATTCACCCTGACACGACAGTCCAAGAGGTCC 1120
          || ||| ||||| || ||| ||||| || || ||| || ||| || || |||||||
Db    121 CATTGGCTCAGGCTACAATCAACTGCCGAATTCACCCTTCGCAGACAGTACATGAGGTCC 180

Qy   1121 TAGAACTCACGAAGAAACATTGTGGGCTGATAAACAGAGACTCCAGTTCCATGTGTTGACTGCCT 1180
          ||| || |||| |||| |||||||||||| || || ||||| || |||||||||||||| ||
Db    181 TAGAACTTGTCAAAGAACACCCGTGGGCTGGTGATGACAGACTCCACCTGCATGTGTTGAGATCCT 240

Qy   1181 TTGACCCCTCCCCGTCAGCCCTTCTGATGACAAGGCTTGGGGTTCACCAGCTGCTCGCGCC 1240
          |||||| || ||| ||||||| ||||||||||||||| ||||||| || |||||
Db    241 TTGAACCCTGCCATCAGCCCTCTGATGACAACAGCATGGGCTTCTACCAGCTGCTTCAAG 300

Qy   1241 AGACCGTACAGGTCCGGTCTTCCCGGAGTCAAATATTACTGCCCAGTTACTTCATTCTTGGCA 1300
          ||||| |||||||| || |||||||| ||| ||| |||||| || ||||||| ||||
Db    301 AGACCATACGATCTGTCTTCCCGGAAGTCGACATCGTCGTCCCGGTATTTGTATTGCCA 360

Qy   1301 ACACGAGACGCCGGATTCTTTTACAAAACCTCACCACTGGCATCTACAGGTTCTACCCCATCT 1360
          || ||| || | |||| |||| |||| ||| ||||||||||||| || |||||||||||
Db    361 ATACGGACACCCGCACCTGGCATCACCAAATGCCATGTTACCCGGTTCAAACCCCCTTC 420

Qy   1361 ACATACAGCCTGAAGACTTCAAACGCATCAATGGAGTGCAACGAGGAAAAATCTCAGTCCAAG 1420
          |||||| || || |||||||||| || ||| |||||||| || |||||| || ||||||
Db    421 CCCTGAACCCTCAGGAGCTTCAGTGGGTCCATGGAATCAATGAGAAAGTTTCCGGTTCAGA 480

Qy   1421 CCTATGAGACCCAAAGTGAAAATTCATCTTTGAGTTGATTCAGAATGCTGACAGAGACCAGG 1480
          || ||| ||||| ||||||||||| ||||||| ||||||| ||||||||| |||| |||
Db    481 ACTACCAGAACCAGGTGAAAGTTCATCCAAAATGCCGACGACACTTACAAAG 540

Qy   1481 AGCCAGTTTCTCACCTGCACAAACTGTGA 1509
          ||||||| |||| || ||| ||| |||
Db    541 AGCCAGTTCCTCATCTGCATGAACTATGA 569
```

RESULT 11
US-10-219-382-2114
; Sequence 2114, Application US/1021382
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Yang, Yonghong
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: Weng, Gezhi
;   APPLICANT: Ma, Yunqing
;   TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
;   FILE REFERENCE: 806A
;   CURRENT APPLICATION NUMBER: US/10/219,382
;   CURRENT FILING DATE: 2002-08-09
;   PRIOR APPLICATION NUMBER: US 60/311,261
;   PRIOR FILING DATE: 2001-08-09
;   PRIOR APPLICATION NUMBER: US 09/488,725
;   PRIOR FILING DATE: 2000-01-21
;   PRIOR APPLICATION NUMBER: US 09/552,317
;   PRIOR FILING DATE: 2000-04-25
;   PRIOR APPLICATION NUMBER: US 09/491,404
;   PRIOR FILING DATE: 2000-01-25
;   PRIOR APPLICATION NUMBER: US 09/519,705
;   PRIOR FILING DATE: 2000-03-07
;   PRIOR APPLICATION NUMBER: US 09/574,454
;   PRIOR FILING DATE: 2000-05-19
;   PRIOR APPLICATION NUMBER: US 09/515,126
;   PRIOR FILING DATE: 2000-02-28
;   PRIOR APPLICATION NUMBER: US 09/560,875
;   PRIOR FILING DATE: 2000-02-03
;   PRIOR APPLICATION NUMBER: US 09/496,914
;   PRIOR FILING DATE: 2000-01-21
;   PRIOR APPLICATION NUMBER: US 09/577,409
;   PRIOR FILING DATE: 2000-05-18
;   PRIOR APPLICATION NUMBER: US 09/515,126
;   PRIOR FILING DATE: 2000-02-28
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 2818
;   SOFTWARE: pt_FLgenes Version 5.0
; SEQ ID NO 2114
;   LENGTH: 362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(362)
;   OTHER INFORMATION: n = a,t,c or g
US-10-219-382-2114

Query Match   16.2%;   Score 245;   DB 10;   Length 362;
Best Local Similarity   79.7%;   Pred. No. 7.5e-61;
Matches   342;   Conservative   0;   Mismatches   0;   Indels   87;   Gaps   1;

```
Qy    1 ATGGCTCAGCGGTGCGTTTGCGTTTGGCTGCGTTGCTGGCCTCTTGTCTGCTGCTGGCCCTGGCTCCCTTGGAGCCCTG   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20 ATGGCTCAGCGGTGCGTTTGCGTTTGGCTGCGTTGCTGGCCCTGGTGGCTGGTCCCTCCTTAGTTTTTCCCT   79

Qy   61 ACCGTCTCCAGATCGATCGAGTGGGCCGAGGAGCGGGAGCCATCAAAAGGGCGTCGGGAATCCCT  120
        ||||||||||||||||||||||||||||||||||||||||||||||
Db   80 ACCGTCTCCAGATCGATCGGGCCCGAGGAGCGGGAGCCATCAAAAGGGCGTCGGGAATCCCT  139

Qy  121 TCTCAGTTGCAGCAAAAGAGGAAACGGCGCTCGCGCCGATGAAAGAGAGCGCGCTGAAAAGGTGCCATCCAG  180
        ||||||||||||||||||||||||||||||||||
Db  140 TCTCAGTTGCAGCAAAAGAGGAACGCGTCGCGCCGATGAAAGAGAGCGCGCTG--------  184

Qy  181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGGTGAGTTCGGA  240
        |||||||||||||||||||||||||||||||
Db  185 -------------AAAGTCTTCCTACAGTGGTGTCAGCACCAGCTTTATTCAGCACATGAAGTC  232

Qy  241 AAATACATTCATTAAAAGTCTTTTCCTACAGTGGTGTCACCACCAGCTTTATTCAGCAGAAGTC  300
        ||||||||||||||||||||||
Db  185 ------------------AAAGTCTTCCTACAGTGGTGTCAGCACCAGCTTTATTCAGCACATGAAGTC  232

Qy  301 GTGGAACGAGTATAGCCACCCTGTTCACTATCTCAAGGCTCGGACACCCAGCTTGCAGCCCTAC  360
        ||||||||||||||||
```

DB 233 GTGGAACGAGTATAGCCACCCTGTTCACTATCTCAAGGCTCGGACACCCTGAAGAAAGGCTCGGAAGAAGGGTCGGGGTGCCCCCA 420
Qy 361 CTGCTGATGGCTCACTTTGATGTGGTGCCTCGCCCCTGAGAAGAAAGGCTCGGAGGGTGCCCCA 420
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 293 CTGCTGATGGCTCACTTTGATGTGGTGCCTCGCCCCTGAGAAGAAAGGCTCGGGGAGGGTGCCCCA 352
Qy 421 TTCTCTGGG 429
        |||||||||
Db 353 TTCTCTGGG 361

RESULT 12
US-10-219-382A-2114
; Sequence 2114, Application US/1021382A
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Yang, Yonghong
;   APPLICANT: Wang, Zhiwei
;   APPLICANT: Weng, Gezhi
;   APPLICANT: Ma, Yunqing
;   TITLE OF INVENTION: Novel Nucleic Acids and Secreted Polypeptides
;   FILE REFERENCE: 806A
;   CURRENT APPLICATION NUMBER: US/10/219,382A
;   CURRENT FILING DATE: 2002-08-09
;   PRIOR APPLICATION NUMBER: US 60/311,261
;   PRIOR FILING DATE: 2001-08-09
;   PRIOR APPLICATION NUMBER: US 09/488,725
;   PRIOR FILING DATE: 2000-01-21
;   PRIOR APPLICATION NUMBER: US 09/552,317
;   PRIOR FILING DATE: 2000-04-25
;   PRIOR APPLICATION NUMBER: US 09/491,404
;   PRIOR FILING DATE: 2000-01-25
;   PRIOR APPLICATION NUMBER: US 09/519,705
;   PRIOR FILING DATE: 2000-03-07
;   PRIOR APPLICATION NUMBER: US 09/574,454
;   PRIOR FILING DATE: 2000-05-19
;   PRIOR APPLICATION NUMBER: US 09/577,409
;   PRIOR FILING DATE: 2000-05-18
;   PRIOR APPLICATION NUMBER: US 09/515,126
;   PRIOR FILING DATE: 2000-02-28
;   PRIOR APPLICATION NUMBER: US 09/560,875
;   PRIOR FILING DATE: 2000-02-03
;   PRIOR APPLICATION NUMBER: US 09/496,914
;   PRIOR FILING DATE: 2000-01-21
;   Remaining Prior Application data removed - See File Wrapper or PALM.
;   NUMBER OF SEQ ID NOS: 2818
;   SOFTWARE: pt_FLgenes Version 5.0
; SEQ ID NO 2114
;   LENGTH: 362
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (1)..(362)
;   OTHER INFORMATION: n = a,t,c or g
US-10-219-382A-2114

Query Match   16.2%;   Score 245;   DB 10;   Length 362;
Best Local Similarity   79.7%;   Pred. No. 7.5e-61;
Matches   342;   Conservative   0;   Mismatches   0;   Indels   87;   Gaps   1;

```
Qy    1 ATGGCTCAGCGGTGCGTTTGCGTTTGGCTGCGTTGCTGGCCTCTTGTCTGCTGCTGGCCCTGGCTCCCTTGGAGCCCTG   60
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   20 ATGGCTCAGCGGTGCGTTTGCGTTTGGCTGCGTTGCTGGCCCTGGTGGCTGGTCCCTCCTTAGTTTTTCCCT   79

Qy   61 ACCGTCTCCAGATCGATCGAGTGGGCCGAGGAGCGGGAGCCATCAAAAGGGCGTCGGGAATCCCT  120
        |||||||||||||||||||||||||||||||||||||
Db   80 ACCGTCTCCAGATCGATCGGGCCCGAGGAGCGGGAGCCATCAAAAGGGCGTCGGGAATCCCT  139

Qy  121 TCTCAGTTGCAGCAGCAAAAGAGGAAACGGCGCTCGCGCCGATGAAAGAGAGCGCGCTG--------  180
        ||||||||||||||||||||||||||||||
Db  140 TCTCAGTTGCAGCAGCAAAAGAGGAACGCGTCGCGCCGATGAAAGAGAGCGCGCTG--------  184
```

Qy   181 ATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 --- 184

Qy   241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCAGCTTTATCCAGCATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   185 ------------------AAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 232

Qy   301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCCTCGGACCCCAGCTTGCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   233 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCCTCGGACCCCAGCTTGCAGCCCTAC 292

Qy   361 CTGCCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAGAGAGGCTGGGAGGTGCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   293 CTGCCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAGAGAGGCTGGGAGGTGCCCCA 352

Qy   421 TTCTCTGGG 429
          |||||||||
Db   353 TTCTCTGGG 361

RESULT 13
US-09-947-907-1617/c
; Sequence 1617, Application US/09947907
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: COLLECTION OF CODING REGION SINGLE
; TITLE OF INVENTION: NUCLEOTIDE POLYMORPHISM (cSNPs) LOCATED ON EACH OF THE HUMAN
; TITLE OF INVENTION: CHROMOSOMES, METHODS OF DETECTION, AND USES THEREOF
; FILE REFERENCE: CL000896
; CURRENT APPLICATION NUMBER: US/09/947,907
; CURRENT FILING DATE: 2003-03-17
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-20
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 21266
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1617
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-947-907-1617

Query Match   15.7%;   Score 236.4; DB 7; Length 601;
Best Local Similarity 96.8%;   Pred. No. 3.1e-58;
Matches 240; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy   248 TTCAPAAAGTCTTTCCTACAGTGGTCAGCAGCCACCAGCTTTATCCACATGAAGTCGTTGGGAAG 307
          | | | ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   498 TGCTTCTAGTGCTTTCCTACAGTGGTCAGCACCAGCTTATCCAGCATGAAGTCGTGGAAG 439

Qy   308 AGTATAGCCACCTGTTCACTATCCAAGCCTCGGACCCCAGCTTGCAGCCCTACCTGCCTGA 367
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   438 AGTATAGCCACCTGTTCACTATCCAAGCCTTGCAGCCTACCTGCCTGA 379

Qy   368 TGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAGAGGCTGGGAGGTGCCCCATTCTCTG 427
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   378 TGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAGAGGCTGGGAGGTGCCCCATTCTCTG 319

Qy   428 GGTTGGAGCCGTGATGGCGTCATCATGAGGTCGGGGCACACTGGCGACGAACAGAAACTCTGTGA 487
          | | | ||||||||||||||||||||||||||||:||||||||||||||||||||
Db   318 GGTTGGAGCCGTGATGGCGCRTCATCTATTGGTCGGGGGCCACACTGGACGACAACAGAAACTCTGTGA 259

Qy   488 TGGCAPTA 495
          ||| |-|
Db   258 TGGCTGA 251

RESULT 14
US-60-452-680-42976

```
Qy    537  CCGAAGATCTTTCTTCATTTC  557
           ||||||||||||||||||||||
Db    181  CCGAAGATCTTTCTTCATTTC  201
```

Search completed: July 1, 2003, 09:11:02
Job time : 1251 secs

us-10-014-896-1.rge

OM nucleic - nucleic search, using sw model

Run on: July 1, 2003, 02:40:21 ; Search time 4024 Seconds

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1 atggctcaagcggtgcgttg.............ctcacctgcacaaactgtga 1509   (without alignments)
                                                                              10913.552 Million cell updates/sec
Scoring table: IDENTITY NUC
                    Gapop 10.0 , Gapext 1.0
Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   GenEmbl:*
        1:    gb_ba:*
        2:    gb_htg:*
        3:    gb_in:*
        4:    gb_om:*
        5:    gb_ov:*
        6:    gb_pat:*
        7:    gb_ph:*
        8:    gb_pl:*
        9:    gb_pri:*
       10:    gb_ro:*
       11:    gb_sts:*
       12:    gb_sy:*
       13:    gb_un:*
       14:    gb_vi:*
       15:    gb_ba:*
       16:    em_fun:*
       17:    em_hum:*
       18:    em_in:*
       19:    em_mu:*
       20:    em_om:*
       21:    em_or:*
       22:    em_ov:*
       23:    em_pat:*
       24:    em_ph:*
       25:    em_pl:*
       26:    em_ro:*
       27:    em_sts:*
       28:    em_un:*
       29:    em_vi:*
       30:    em_htg_hum:*
       31:    em_htg_inv:*
       32:    em_htg_other:*
       33:    em_htg_mus:*
       34:    em_htg_pln:*
       35:    em_htg_rod:*
       36:    em_htg_mam:*
       37:    em_htg_vrt:*
       38:    em_sy:*
       39:    em_htgo_hum:*
       40:    em_htgo_mus:*
       41:    em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB ID	Description
1		1373.4	91.0	2101 9	AK057131	Homo sapi
2		1310.2	1422	6	AX319870	Sequence
3		1310.2	86.8	1779 6	AX480926	Sequence
4		1310.2	86.8	3071 10	BC025830	Mus muscu
5		673.6	44.6	188119 2	AC119673	Homo sapi
6	c	235.2	12.3	185081 2	AC079937	Mus muscu
7		186	11.5	213050 2	AL359702	Mus muscu
8		173	12.3	156313 2	AL359702	Mus muscu
9		170.4	11.5	171999 2	AL646079	Rattus no
10	c	131.6	8.7	156313 2	AC121225	Rattus no
11		125.6	8.3	156313 2	AC121225	Rattus no
12		120.2	7.3	134617 2	AC115001	Homo sapi
13	c	110.2	7.0	5664 1	AC115001	Mus muscu
14		106.2	6.6	86564 1	AE005748	Caulobact
15		93.2	6.2	171999 2	AC121225	Uncultur
16		59.2	3.9	188119 9	AC119673	Homo sapi
17		56	3.7	33560 1	AC119673	Rattus no
18		51.4	3.4	1469 10	SPAC24C9	S.pombe chr
19		51	3.4	786 1	SC9C7	Streptomy
20		50.8	3.4	1221 6	AX157288	Sequence
21		50.8	3.4	1304 4	E04019	DNA sequenc
22		50.8	3.4	1341 4	SCAC1	S.scrofa mR
23		49.4	3.3	51 1	PIGACY1	Porcine mRN
24		49.4	3.3	51 1	AX157285	Sequence
25		49.2	3.3	1224 6	AX157287	Sequence
26		49.2	3.3	1224 6	E04020	cDNA sequen
27		49.2	3.3	1224 6	E07281	genomic DNA
28		49.2	3.3	1224 6	I18856	Sequence 3
29		49.2	3.3	1224 6	I32240	Sequence 3
30		49.2	3.3	1388 9	HUMACY1	Human N-acy
31		49.2	3.3	1403 9	D14524	Human mRNA
32		49.2	3.3	1415 9	HUMACYPRO	Human amino
33	c	49.2	3.3	1416 9	L07548	Human amino
34		49.2	3.3	1423 9	BC003023	Homo sapi
35		49.2	3.2	1432 9	BC014112	Homo sapi
36		48.8	3.2	10861 1	AE005988	Caulobact
37		47.8	3.2	51 1	AX157286	Sequence
38		47.6	3.1	12249 1	AE002039	Deinococc
39		46.2	3.1	381 1	AX307434	Sequence
40	c	46.2	3.1	7066 1	SC5C5	Streptomy
41		45.6	3.0	147215 8	ATF20D10	Arabidops
42		45.6	3.0	199789 8	ATCHRIV89	Arabidops
43		44.8	3.0	663 6	AX364389	Sequence
44		44.8	3.0	787 6	AX364030	Sequence
44		44.8	3.0	799 6	AX364183	Sequence
45		44.8	3.0	1542 8	AY074847	Arabidops

ALIGNMENTS

RESULT 1
AK057131

LOCUS   AK057131   2101 bp   mRNA   linear   PRI 01-AUG-2002
DEFINITION   Homo sapiens cDNA FLJ32569 fis, clone SPLEN2000134, weakly similar
             to CARBOXYPEPTIDASE S PRECURSOR (EC 3.4.17.4).
ACCESSION   AK057131
VERSION   AK057131.1 GI:16552718
KEYWORDS   oligo capping; fis (full insert sequence);
SOURCE   Homo sapiens spleen cDNA to mRNA, clone_lib:SPLEN2
             clone:SPLEN2000134.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kawakami,B., Sugiyama,A., Takemoto,M., Sugiyama,T., Irie,R.,

TITLE   Isogai,T., Otsuki,T. and Sugiyama,T.

REFERENCE   2 (bases 1 to 2101)
AUTHORS   Otsuki,T., Sato,H., Ota,T., Wakamatsu,A., Ishii,S., Yamamoto,J.,
            Isono,Y., Kawai-Hio,Y., Saito,K., Nishikawa,T., Kimura,K.,
            Yamashita,H., Matsuo,K., Nakamura,Y., Sekine,M., Kikuchi,H.,
            Kanda,K., Wagatsuma,M., Murakawa,K., Kanehori,K.,
            Takahashi-Fujii,A., Oshima,A., Suzuki,Y., Sugano,S., Nagahari,K.,
            Masuho,Y., Nagai,K. and Isogai,T.
TITLE   NEDO human cDNA sequencing project
JOURNAL   Unpublished

REFERENCE   2 (bases 1 to 2101)
AUTHORS   Isogai,T., Otsuki,T. and Sugiyama,T.
TITLE   Direct Submission
JOURNAL   Submitted (24-OCT-2001) Takao Isogai, Helix Research Institute,
            Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
            (E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT   NEDO human cDNA sequencing project supported by Ministry of
            Economy, Trade and Industry of Japan; cDNA full insert sequencing:
            Research Association for Biotechnology (RAB); cDNA library
            construction: Helix Research Institute (HRI) (supported by Japan
            Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB,
            HRI, and Biotechnology Center, National Institute of Technology and
            Evaluation; clone selection for full insert sequencing: RAB and
            HRI.

FEATURES   Location/Qualifiers
     source   1..2101
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="SPLEN2000134"
                     /clone_lib="SPLEN2"
                     /tissue_type="spleen"
                     /note="cloning vector: pME18SFL3"
     CDS   61..1146
                     /codon_start=1
                     /note="unnamed protein product"
                     /protein_id="BAB71368.1"
                     /db_xref="GI:16552719"
                     /translation="MAQRCVCVLALVAMLLLVPFTVSRSMGPRSGEHQRASRIPSQFS
                     KEERVAMKEALKGAIQIPVTFSSEKSNTIALAEFGKYIHKVFPPTVVSTSFIQHEVVE
                     EYSHLFTIQGSDPSLQPYILMAHFDVPAPEEGWEVPPSGLERDGVIYGWGTLDDKN
                     SVMAILQALELLLIRKYIPRRSFFISLGHDEESSGTGAQRISALIQSRGVQLAFIVDE
                     GGFILDDFPINFKRPIALIAVSEKGSMNLMLQVNMTSGHSSARPKETSIGILAAAVSR
                     LEQTPMPIIPGSGTVVTVLQQLANEVYGEKSLNQCNNQDHHGTHHIQSRGQVQCHPPS
                     GPGHSQLPDSPWTDSPRGPRTHEEHCG"

BASE COUNT   507 a   579 c   477 g   538 t
ORIGIN

Query Match 91.0%; Score 1373.4; DB 9; Length 2101;
Best Local Similarity 95.8%; Pred. No. 0; Mismatches 1; Indels 62; Gaps 1;
Matches 1446; Conservative 0; Mismatches 1; Indels 62; Gaps 1;

Qy   1 ATGGCTCAGCCGGTGCCTTTGCGTTGGCTGGCCCCTTGGTGGCTCTCTCCTTAGTTTCCCT   60
            |||||||| ||||||||||||||| ||||| ||| ||||||||||||||||||||||||||
Db   61 ATGGCTCAGCCGGTGCCTTTGCGTTTGGCTGGCCCCTTGGTGGCTCTCCTAGTTTTCCCT   120

Qy   61 ACGGTCTCCAGATACGATGGGCCCCGAGGAGCGGGGAGCCATCAAAGGGCGTCGGAATCCCT 120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 ACGGTCTCCAGATACGATGGGCCCCGAGGAGCGGGGAGCCATCAAAGGGCGTCGGAATCCCT 180

Qy   121 TCTCAGTTCAGCAAAGAAGAGGAACGCGTCGCCGATGAAAGAAGGGTGCCATCCAG 180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 TCTCAGTTCAGCAAAGAAGAGGAACGCGTCGCCGATGAAAGAAGGGTGCCATCCAG 240

Qy   181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCGCCCTGGCCTGAGTTCGGA 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db   241 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCGCCCTGGCCTGAGTTCTTT 300

Qy   241 AAATACAATCATAAAAGTCTTTCCTACAGTGGTCAGCAGCATGAAGTC 300
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   301 AAATACAATCATAAAAGTCTTTCCTACAGTGGTCAGCAGCATGAAGTC 360

Qy   301 GTGGAAGAGTATAGCCACCACTGTTCACTATCCAAAGGCTCGGAC 360
            |||||||||||||||||||||||||||||||||||||||||||||
Db   361 GTGGAAGAGTATAGCCACCACTGTTCACTATCCAAAGGCTCGGAC 420

Qy   361 CTGCTGATGGCTCACTTTGATGTGGTCGCCCCCTGCCCCCA 420
            ||||||||||||||||||||||||||||||||||||||||||
Db   421 CTGCTGATGGCTCACTTTGATGTGGTCGCCCACTGACACAGAAC 480

Db   481 TTCTTCTCGGGTTGGAGCCGTGATGGCGTCATCTATGGTGG 540

Db   541 TCTGTGATGGCATTACTGCAGCCCTTGGAGCTC 600

Qy   541 AGATCTTCTTCATTTCTCTGCGCCAT 600
            ||||||||||||||||||||||||||
Db   601 AGARCTTTCTTCATTTCTCTGCGCCAT 660

Qy   601 ATCTCAGCCCTCTACAGTCAAGGGCGTCCAGCTGCC 660
            ||||||||||||||||||||||||||||||||||||
Db   661 ARTCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCT 720

Qy   661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGA 720
            ||||||||||||||||||||||||||||||||||
Db   721 TTCATCTTGGATGATTTCATTCCTAACTTCAAGA 780

Qy   721 GAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGG 780
            |||||||||||||||||||||||||||||||||||||||||||||||
Db   781 GAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGG 840

Qy   781 CCTCCCAAAAGGAGACAAAGCAGTTGGCATCTT 840
            |||||||||||||||||||||||||||||||||
Db   841 CCTCCCAAAAGGAGACAAAGCAGTTGGCATCTT 900

Qy   841 CCAATGCCTATCATATTTGGAAGCGGGAACAGTGGTGACTGTATTG 900
            ||||||||||||||||||||||||||||||||||||||||||||||
Db   901 CCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTG 960

Qy   901 GAGTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGC 960
            ||||||||||||||||||||||||||||||||||||||||||
Db   901 GAGTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGC 960

Qy   961 G-- 961
Db   961 GAGGTTTATGGAGAGAAAATCCCTTAACCAATG 1020

Qy   961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGC 1020
            ||||||||||||||||||||||||||||||||||||
Db   962 ---AGGTTTATGGAGAGAAAATCCCTTAACCAATGC 1018

Qy   1021 ACCATATTCAAAACGACAGGGGTCAAGTTCATCCC 1080
            |||||||||||||||||||||||||||||||||||
Db   1019 ACCATATTCAAAACGACAGGGGTCAAGTTCATCCC 1078

Qy   1081 AACTTCCGGATTCACCCTGGACAAGAC 1140
            |||||||||||||||||||||||||||
Db   1079 AACTTCCGGATTCACCCTGGACAAGAC 1138

Qy   1141 GTGGCTGATAACAGAGTGCCAGTTCCA 1200
            |||||||||||||||||||||||||||
Db   1139 GTGGCTGATAAACAGAGTGCCAGTTCCA 1198

Qy   1201 CCTTCTGATGACAAGGGCTACCAGTCC 1260
            |||||||||||||||||||||||||||
Db   1199 CCTTCTGATGACAAGGGCTACCAGTCC 1258

Qy   1261 CCGGAAGTGCAAATATTACTGCCCCAGTTAC 1320
            |||||||||||||||||||||||||||||||
Db   1259 CCGGAAGTGCAAATATTACTGCCCCAGTTAC 1318

Qy   1321 ACAAACCTCACCACTGGCATCACAGTTC 1380
            ||||||||||||||||||||||||||||
Db   1319 ACAAACCTCACCACTGGCATCACAGTTC 1378

Qy   1381 AAACGCATCCATGGAGTCAACGAGAAA 1440
            |||||||||||||||||||||||||||
Db   1379 AAACGCATCCATGGAGTCAACGAGAAA 1438

Qy   1441 TTCATCTTTGAGTTGATTCAGAATGC 1500
            ||||||||||||||||||||||||||
Db   1441 TTCATCTTTGAGTTGATTCAGAATGC 1500

```
RESULT 2
LOCUS       AX319870
DEFINITION  Sequence 34 from Patent WO0183782.
ACCESSION   AX319870
VERSION     AX319870.1  GI:17901460
KEYWORDS
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Plowman,G.D., Whyte,D., Sudarsanam,S., Manning,G., Caenepeel,S. and
            Payne,V.
  TITLE     Novel proteases
  JOURNAL   Patent: WO 0183782-A 34 08-NOV-2001;
            Sugen, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..1422
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
BASE COUNT      352 a    390 c    350 g    330 t
ORIGIN

Query Match         86.88%;  Score 1310.2;  DB 6;  Length 1422;
Best Local Similarity   94.0%;  Pred. No. 0;  Mismatches 3;  Indels 87;  Gaps 2;
Matches 1419;  Conservative   0;  Mismatches   3;  Indels  87;  Gaps   2;
```

```
Db   1439 TTCATCTTTGAGTTGATTCAGAAATCCTGACACAGACCAGGAGCCAGTTGTTCTCACCTGCAC 1498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   1501 AAACTGTGA 1509
          |||||||||
Db   1499 AAACTGTGA 1507
```

```
Qy      1 ATGGCTCAGCGGTGCGTTGCGTGCGTCGGCCCTGGGTGCTANGCCTCGGTANGCTCGCTGGT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 ATGGCTCAGCGGTGCGTTGCGTGCGTCGGCCCTGGGTGCCTANGCCTCGCTCCTAGTTTTCCCT 60

Qy     61 ACCGTCTCCAGATCGATCGGGGCCCCGAGAGGAGCCANCAAAGGCGGTCGCGGAATCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 ACCGTCTCCAGATCGATCGGGGCCCCGAGAGGAGCCGATATCAAAGGCGGTCGCGGAATCCCT 120

Qy    121 TCTCAGTTCAGCAAAGAGAGAGAACCCGTCCGGATGAAAAGAGGCCTTGAAAGGTGCCATCCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 TCTCAGTTCAGCAAAGAGAGGAACCGGTCCGGATGAAAAGAGGCACTGAAAGGTGCCATCCAG 180

Qy    181 ATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAATACTACAGCGCCCTGGCTGAGTTCCGA 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 ATTCCAACAGTGACTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGCTGAGTTCCGGA 240

Qy    241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACACCACCAGCTTTATCCAGCATGAAGTC 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 AAATACATTCGTAAAGTCTTTCCTACAGTGGTCAGCACACCAGCTTTATCCAGCATGAAGTC 300

Qy    301 GTGGAAGAGTATAGGCACCACCTGTTCACTATCCAAGGCTCCGACCCCAGCGTTGCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGAAGAGTATAGGCACCACCTGTTCACTATCCAAGGCTCCGACCCCAGCGTTGCAGCCCTAC 360

Qy    361 CTGCTGATGGCTCACTTTGATGTGTGCCTGCCCCTGAGAAGGCTGTGGGAGTGCCCCCA 420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 CTGCTGATGGCTCACTTTGATGTGTGCCTGCCCCTGAGAAGGCTGTGGGAGTGCCCCCA 420

Qy    421 TTCTCTCTGGGTTGGAGCGTGATGGGGGTCATCTATGGTGCGGGACAACAAGAAC 480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 TTCTCTCTGGGTTGGAGCGTGATGGGGGTCATCTATGGTGCGGGACAACAAGAAC 480

Qy    481 TCTGTGATGGCATTACTGCAGGCCGTTGGGAGCTCCTGCTTGAGGAAGTACATCCCCCGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 TCTGTGATGGCATTACTGCAGGCCGTTGGGAGCTCCTGCTTGAGGAAGTACATCCCCCGA 540

Qy    541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGGACTCATCAGGGACAGGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGGACTCATCAGGGACAGGGGCTCAGAGG 600

Db    541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGGACTCATCAGGGACAGGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy    601 ATCTCAGCCCTGCTACAGCAGTCAAAGGGGCGGTCCAGGCTAGCCTTCATTGTGGACGACAGG 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601 ATCTCAGCCCTGCTACAGCAGTCAAAGGGGCGGTCCAGGCTAGCCTTCATTGTGGACGACAGG 660

Qy    661 TTCATCTTTGAGTGATTCAGAAATCCTAACTTCAAGAAGCCCATCGCCCTTGATTGCAGTCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 TTCATCTTTGAGTGATTCAGAAATCCTAACTTCAAGAAGCCCATCGCCCTTGATTGCAGTCTCA 720

Qy    721 GAGAAGGGGTTCCATGAACCTCATTGCTGCAGTAAACATGACTTCAGGCCACTCTTCAGCT 780
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 GAGAAGGGGTTCCATGAACCTCATTGCTGCAGTAAACATGACTTCAGGCCACTCTTCAGCT 780

Qy    781 CCTCCAAAGGAGGACAAAGCCATTGGGCATCCTTCCAACTTCAAAGAAGCCCATCGCCCTTG 840
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 CCTCCAAAGGAGGACAAAGCCATTGGGCATCCTTCCAACTTCAAAGAAGCCCATCGCCCTTG 840

Qy    841 CCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTTATTGCAGCAACTGGCAAAT 900
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTTATTGCAGCAACTGGCAAAT 900

Qy    901 GAGTTTCCCTTCCCTGTGCAAATATAAATCCTGACCAACCCATGCTATTTGAACCACTTATA 960
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||--------
Db    901 G------------------------------------------------------------------- 901

Qy    961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAATAAATCAGGACCAACGGCGCACTC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    902 ---AGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAATAAATCAGGACCAACGGCGCACTC 958

Qy   1021 ACCATAATTCAAAGCAGGGGGGTCAAGTTCAAATGTCATCCCCCAGTGGGCCCACGGCCCACAGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||----------------
Db    959 ACCATAATTCAAAGCAGG-------------------------------GTGGGCCCAGCCCACAGTC 993

Qy   1081 AACTTCCCGGATTCACCCTGGACAGACAGTCCAAAGAGGTCCTAGAACTCACGAAGAACATT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    994 AACTTCCCGGATTCACCCTGGACAGACAGTCCAAAGAGGTCCTAGAACTCACGAAGAACATT 1053

Qy   1141 GTGGCTGATAAACAGAGTCCAGTGTTCCATGTGCTTTGCCTTTGACCCCGTCCCCGTCAGC 1200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1054 GTGGCTGATAAACAGAGTCCAGTGTTCCATGTGCTTTGCCTTTGACCCCGTCCCCGTCAGC 1113

Qy   1201 CCTTCTGATGACAAGGCCTTGGGCTACCAGCGCTGCTCCGCCAGACCGTACAGTCCGTCTTC 1260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1114 CCTTCTGATGACAAGGCCTTGGGCTACCAGCGCTGCTCCGCCAGACCGTACAGTCCGTCTTC 1173

Qy   1261 CCGGAAGTCAATATTACTGCCCCAGTACTTCTTCTTTATTGCAACACACAGAGCAGCCGGATTCTT 1320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1174 CCGGAAGTCAATATTACTGCCCCAGTACTTCTTCTTTATTGCAACACACAGAGCAGCCGGATTCTT 1233

Qy   1321 ACAAAACCTCACCCACTGGCATCTACAGGGTTGCTACTTCATTACAGACATGAAGACTTC 1380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1234 ACAAAACCTCACCCACTGGCATCTACAGGGTTGCTACTTCATTACAGACATGAAGACTTC 1293

Qy   1381 AAAGGCATCCATGGAGTCAACGAGAAAAATCACTCAAAGGCCTAATGAGACCAAGTGAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1294 AAAGGCATCCATGGAGTCAACGAGAAAAATCACTCAAAGGCCTAATGAGACCAAGTGAA 1353

Qy   1441 TTCATCTTTGAGTTGATTCAGAAATCCTGACACAGACCAGGAGCCAGTTGTTCTCACCTGCAC 1500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1354 TTCATCTTTGAGTTGATTCAGAAATCCTGACACAGACCAGGAGCCAGTTGTTCTCACCTGCAC 1413

Qy   1501 AAACTGTGA 1509
          |||||||||
Db   1414 AAACTGTGA 1422
```

```
RESULT 3
LOCUS       AX480926
DEFINITION  Sequence 22 from Patent WO0246383.
ACCESSION   AX480926
```

```
VERSION      AX480926.1  GI:22217565
KEYWORDS     human.
SOURCE       Homo sapiens
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1
  AUTHORS    Yue,H., Azimzai,Y., Kallick,D.A., Baughn,M.R., Griffin,J.A.,
             Swarnakar,A., Lal,P.G., Walia,A.J., Hafalia,A.J., Gandhi,A.R.,
             Au-Young,J., Elliott,V.S., Ramkumar,J., Thangavelu,K., Lu,Y.,
             Warren,B.A., Lu,D.A., Lee,E.A., Tribouley,C.M., Arvizu,C.,
             Delegeane,A.M., Yao,M.G., Khan,F.A. and Sanjanwala,M.M.
  TITLE      Protein modification and maintenance molecules
  JOURNAL    Patent: WO 024638-A 22 13-JUN-2002;
             Incyte Genomics Inc (US)
FEATURES             Location/Qualifiers
     source          1..1779
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /note="Incyte ID NO: 7481673CB1"
BASE COUNT    441 a    488 c    440 g    410 t
ORIGIN

Query Match             86.8%;  Score 1310.2;  DB 6;  Length 1779;
Best Local Similarity   94.0%;  Pred. No. 0;
Matches 1419;  Conservative    0;  Mismatches   3;  Indels   87;  Gaps   2;

Qy      1 ATGGCCTCAGCGCGTGCGGTGCGTCGTTGCGTCGGTGCGCCTGTGGTCGGCGTATGCGGG 60
          ||||||||||||||| ||||| ||| ||| |||||||||||||||||||||||||| ||||
Db    109 ATGGCCTCAGCGCGGGTGCGGTTGCGGTTGCCTCGCTGGTGCGCCTGTGGTCGGCGTTTTTCCCT 168

Qy     61 ACCGGTCTCCAGATCGAATCGGGGCCCGAGGAGCGGGAGCATCAAAGGGCCGTCGCGAATCCCT 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    169 ACCGGTCTCCAGATCGAATCGGGGCCCGAGGAGCGGGAGTATCAAAGGGCCGTCGCGAATCCCT 228

Qy    121 TCTCAGTTCACGAAAGAGAGGAACGCGTCTCGCCATGAAAGAGGCGCTGAAAGGTGCCATCCAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    229 TCTCAGTTCACGAAAGAGAGGAACGCGTCGCCATGAAAGAGGCACTGAAAGGTGCCATCCAG 288

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCCTGGCTGACTTCGGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    289 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCCTGGCTGACTTCGGA 348

Qy    241 AAATACATTCATAAAGTCTTTCCTACACAGTGGCTCAGGCTTTATCCAGCATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    349 AAATACATTCGTAAAGTCTTTCCTACACAGTGGCTCAGGCTTTATCCAGCATGAAGTC 408

Qy    301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    409 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 468

Qy    361 CTGCTGATGGGCTCACTTTGATGTGGTGCCTCGGAGGTGGGGAGGTGCCCCCA 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    469 CTGCTGATGGGCTCACTTTGATGTGGTGCCTCGGAGGTGGGGAGGTGCCCCCA 528

Qy    421 TTCTCTCTGGGTGTTGGAGGCGTGCATCTTTGGTCGCGGGGGCACGCAAGAAC 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    529 TTCTCTCTGGGTTGGAGCCTGATCGGCGTCATCTATTGGTCGCGGGGGCACGCAAGAAC 588

Qy    481 TCTGTGATGGCATTACTGCAGGCTCCTCCTGCTCATAAGGAAGTAACATCCCCGA 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    589 TCTGTGATGGCATTACTGCAGGCTCCTCCTGCTCATAAGGAAGTAACATCCCCGA 648

Qy    541 AGATCTTTCTTCATTCTCTGGGGCCATGAGATGAGGACACAGGGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    649 AGATCTTTCTTCATTCTCTGGGGCCATGAGATGAGGACACAGGGCTCAGAGG 708

Qy    601 ATCTCAGCCCGTCTACAGTCAAGGGGCGTCCAGCTAGCCTTCATCAGGGACAGGGGGC 660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    709 ATCTCAGCCCGTCTACAGTCAAGGGGCGTCCAGCTAGCCTTCATTGTGACAGAGGGGGC 768

Qy    661 TTCATCTTGGATGATTTCATTCCTAAGAAGCCCATCCGCCCTTGATTGCAGTCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    769 TTCATCTTGGATGATTTCATTCCTAACTTCAAGAGGAAGCCCATCCGCCCTTGATTGCAGTCTCA 828

Qy    721 GAGAAGGGTTCCATGAACCCCATGCTGCGAAGTAAACATGACTTCAGGCCACTCTTCAGCT 780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    829 GAGAAGGGTTCCATGAACCCCATGCTGCGAAGTAAACATGACTTCAGGCCACTCTTCAGCT 888

Qy    781 CCTCCAAAGGAGACAAGCAATTGGCATCCTTGCCAGCTGCGTGTCAGCCGGATTGGAGCAGACA 840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    889 CCTCCAAAGGAGACAAGCAATTGGCATCCTTGCCAGCTGCGTGTCAGCCGGATTGGAGCAGACA 948

Qy    841 CCAATGCCTATCATATTTGGAAGCGGGGACAGTGGACTGGGTCTATTGCACGCAACTGGGAAAAT 900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    949 CCAATGCCTATCATATTTGGAAGCGGGGACAGTGGACTGGGTCTATTGCACGCAACTGGGAAAT 1008

Qy    901 GAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCCATGGCTATTTGAACCACTTATA 960
          |                                                           |
Db   1009 G-                                                           1009

Qy    961 AGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAAATAATCAGGACCACCACGCACTC 1020
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1010 ---AGGTTTATGGAGAGAAAACCCTTAACCAATGCAAATAATCAGGACCACCACGCACTC 1066

Qy   1021 ACCATATTCAAAGCACAGGGGTCAAGTTCAATGTTCATCCCCCAGTGGGCCCACAGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||   ||||||||||
Db   1067 ACCATATTCAAAGCACAGG-----------------------------GTGGGCCCACAGTC 1101

Qy   1081 AACTTCCGGATTCACCCTGGACACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1102 AACTTCCGGATTCACCCTGGACACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACATT 1161

Qy   1141 GTGGGCTGATAACAGAGTCCACTGCCAGTTCCAGTGTTGAGTGGTTTTGGTGAGTGCTTTGAC 1200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1162 GTGGGCTGATAACAGAGTCCAGTCCAAGTGTTGAGTGGTTTGAGTGCTTTTGACCCCCGTCAGC 1221

Qy   1201 CCTTCTGATGACAAGGCGTACCAGCGTGCTCCGGCCGACTGCAGTCCGTCTTC 1260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1222 CCTTCTGATGACAAGGCCCTTGGGGTTCACCAGCGTACCAGTCCGGTCTTC 1281

Qy   1261 CCGGAAGTCAATAATTACTGCCCCCAGTTACTTCTATTGGCAACACAGAAGCCATTCTTT 1320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1282 CCGGAAGTCAATATTACTGCCCCCAGTTACTTCTATTGGCAACACAGAAGCCATTCTTT 1341

Qy   1321 ACAAAACCTCACCACTGCATCTCACAGTTCTACCCCATCCTGAAGACTTC 1380
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1342 ACAAACCTCACCACTGCCATCCTACAGTTCTCTACTTCTATCCTGAAGACTTC 1401

Qy   1381 AAACGCATCCATGGAGTCAAGAGAAAAATCTCAGTCCCATCTACAACACCAAGTGAAA 1440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1402 AAACGCATCCATGGAGTCAAGAGAAAATCTCAGTCTCTATGAGACCAAGTGAAA 1461

Qy   1441 TTCATCTTTGAGTTGATTCAGAATGGCAGACAAGACCAGTTGTTCTCACCTGCAC 1500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1462 TTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGTTTCTCACCTGCAC 1521

Qy   1501 AAACTGTGA 1509
          |||||||||
Db   1522 AAACTGTGA 1530
```

RESULT 4
```
BC025830
LOCUS       BC025830                3071 bp    mRNA    linear   ROD 07-AUG-2002
DEFINITION  Mus musculus, clone IMAGE:5124369, mRNA, partial cds.
ACCESSION   BC025830
VERSION     BC025830.1  GI:19387946
KEYWORDS
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3071)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
```

```
JOURNAL   Submitted (11-MAR-2002) National Institutes of Health, Mammalian
          Gene Collection (MGC), Cancer Genomics Office, National Cancer
          Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
          USA
REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT   Contact: MGC help desk
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Jeffrey E. Green, M.D.
          Contact: MGC help desk
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: the I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Genome Sequence Centre,
          BC Cancer Agency, Vancouver, BC, Canada
          info@bcgsc.bc.ca
          Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
          Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
          Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
          Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
          Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline
          Schein, Duane Smailus, Michael Smith, Lorraine Spence, Jeff Stott,
          Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,
          George Yang, Scott Zuyderduyn, Marco Marra.

          Clone distribution: MGC clone distribution information can be found
          through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
          Series: IRAK Plate: 57 Row: 1 Column: 18
          This clone was selected for full length sequencing because it
          passed the following selection criteria: Hexamer frequency ORF
          analysis.
FEATURES          Location/Qualifiers
     source        1..3071
                   /organism="Mus musculus"
                   /db_xref="taxon:10090"
                   /map="FVB/N"
                   /clone="IMAGE:5124369"
                   /tissue_type="Liver, normal. 5 month old male mouse."
                   /clone_lib="NCI_CGAP_Li9"
                   /lab_host="DH10B"
                   /note="Vector: pCMV-SPORT6"
     CDS           <1..1025
                   /codon_start=3
                   /product="Unknown (protein for IMAGE:5124369)"
                   /protein_id="AAH25830.1"
                   /db_xref="GI:19387947"
                   /translation="AILHALELLLIRNYSPKRSFFIALGHDEEVSGEKGAQKISALLQ
                   ARGVLAFLVDEGSPILEGFIPNLEKPVAMISVTEKGALDLMLQVNMTPGHSSAPFKE
                   TSIGILSAAVSRLEQTPMPNMFGGGPLKKTMKILANEFSPINIVLRNLWLFHPIVSR
                   IMERNPITNALVRTTFALTMFNAGIKVNVIPPLAQATINCRHIPSQTVHEVLELVKNT
                   VADDRVQLHYLRSFEPLPISPSDDQAMYQLLQEYIRSVREVPEVDIVVPGICIANFDTR
                   HYANTINGMRFNFLPLNPQDFSGVHIGINEKVSVQNYQNQVKFIFEFIQNADTYKEPV
                   PHLHEL"
BASE COUNT      783 a    808 c    696 g    784 t
ORIGIN
```

Query Match 44.6%; Score 673.6; DB 10; Length 3071;
Best Local Similarity 79.38%; Pred. No. 1.5e-183;
Matches 812; Conservative 0; Mismatches 209; Indels 3; Gaps 1;

```
Qy   489 GGCATTACTGCCAGGCCGCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGGAAGATCTT 548
         ||| |||||| ||||||| ||||||||| ||  |||||||||| ||||| ||||| |||||||
Db     2 GGGGATCCTGCATGCGTTTGGGAGCTCCGTGTTGATCAGGAAACTACAGCCCCAAAGAGATCTTT 61

Qy   549 CTTCATTTCTCTGGGCCAGGCCATGAGATGAGGAGTCATCAGGG---ACAGGGCGTCAGAGGATCTC 605
         |||||| || ||||||||| || |||||||||||||||||| ||   | |||||||||||||||||||
Db    62 CTTCATTGCTTGGGCCATGATGAGGAGGAGTGTCCGGGGGAAAAAGGGGCGTCAGAGATCTC 121

Qy   606 AGCCCTGCTACACTCAAGGGCGGGCTCCACTCAGCCTTTCATTGTGGCACGAGAGGGGGCTTCAT 665
         || ||||||||| ||||||| |||||| |||||||||||| ||||| ||| ||||||| |||||
Db   122 AGCACTCTTACACGGGGCGGTTGCCAGCTAGCCTTCCTTGTGGGGATGGAGGGAGCTTCAT 181

Qy   666 CTTGGATGATTTCATTTCCTAACTTCAAGCAGCCCATGCCGCTTTCTTGCAGTCTTCAGCGAA 725
         |||||| |||||| |||  |||| || ||||||||||||  ||| |||| |||||||||| |||
Db   182 CTTGGAAGGCTTCATTCAAACCTCGAGAAGCCAGTTGCCATGATTTCAGTCACTGAGAA 241

Qy   726 GGGTTCCATGAAACCTCATGCTGCGAAGTAAACATGACTTCAGGCCCACTCTTCAGCTCCTCC 785
         |||| ||  || |||||||||||| |||||||||||||||||||||| ||||||||||||||
Db   242 GGGTGCCCTTGACCCTCATGCTGCGAAGTAAACATGACTTCAGGCCCACTCTTCAGCTCCCC 301

Qy   786 AAAGGAGACAAAGCAATTGGCCATCCTTGCCAGCGTGCCTGCTGTCAGGAGGACCAACTGCCAAATGACCAAT 845
         |||||||||||||| ||||| |||||||| |||||||| ||||| |||||||||||||||||||||||||||||
Db   302 AAAGGACACAAAGCAATTGGCCATCCTTCTTCGCCGTGCCTGGTGTCAGGAGGACCAAT 361

Qy   846 GCCTATCATATTTGAAGCGCGGGACAAGTGGGTGACTGTTATTCCGCAACTGCCAAATGAGTT 905
         ||  |||||||||||||| ||||||||||||||||||||||| |||| ||||||| ||||||||
Db   362 GCCGAATATGTTTGGAGGAGGGCCATTGAAGACAAATGAAGCTTACTGGCAAATGAGTT 421

Qy   906 TCCCTTCCCTGTCAATATAATACCTGAGCAACCCATGGCTATTTGAACCACTTTATAAGCAG 965
         || ||| | |||||| || |||| |||| |||| ||||| | ||||| |||||||| ||||||
Db   422 TTCCTTCCCTATCAATATTAGTCTTGAGAAAACCTGTGGGCTATTTCATCCCATTGTGAGCAG 481

Qy   966 GTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGAGACCACCACGGCCGCCACCAT 1025
         ||||| ||||||||| ||||| |||| ||||||||||| ||||||| |||||||||| ||||||
Db   482 GATTAATGGAGAGGAAACCCATCAAACAAATGCCCTGGTCCGGACCTCCAGCGCCCTCACCAT 541

Qy  1026 ATTCAAAGCACAGGGGGTCAAGTTCCATAGTGTCATCTCATCCCCCAGTGGCCCCAGGCCACAGTGACAACTT 1085
         | ||||| |||||||| || ||||||  ||||||  || ||  | |||||||||||||||| |||| |||||
Db   542 GTTCAATGCAGGAAATCAAGGTGAAATGTCATCCCTCCATTGGCTCAGGCTACAAATCAAACTG 601

Qy  1086 CCGGATTCACCCTGGACAAGCAGTCCAAGAGGTCCTAGAACTCAGGAAGAACATTGTGGC 1145
         ||| ||||| |||||||||||| |||| |||| ||||||||| ||||||||||||||||||
Db   602 CCGAATTCACCCTTCGGCACAGTACATGAGGTCCTAGAACTTGTCAAGAACACCGTGGC 661

Qy  1146 TGATAACAGAGTCCAGTTCCATAGTGTGTTGAGTGGCCTTGAGCCCGGTCCAGCCCTTC 1205
         |||| ||||| ||||||||| |||||||||||||| | |||||||||||||||||| |||
Db   662 TGATGACAGAGTCCAGCTCAGTCCATAGTGTTGAGTGGCCTTGAACCCCTGCCCATCAGCCCTC 721

Qy  1206 TGATGACAAAGGCCTTGGGCGTTCCGGCAGCGTGCTCCGCGCAGACCCTGTACGAGTCGGTGCTCTTCCCGGA 1265
         ||||| |||||||| |||||||| ||| |||||||| |||||||||||| ||||||||||| |||||||||||||
Db   722 TGATGACCAGGGCCATGGGCTACCAGCTACCGGTTCAACCCCTTCCCGGA 781

Qy  1266 AGTCAATATTACTGCCCCAGTTACTTCTATTGGGCAACAGCAGCGGATTCTTTACAAA 1325
         || |||||||  |||||| ||||| ||  |||||| ||||||||| |||||||| |||
Db   782 AGTCGACATCGTGGTCCCCGGTATTTGTGATTGCCAATACGGACACCCGACTATGCCAA 841

Qy  1326 CCTCACCACTGGGCCATCTACAGGTTCTACCCCCATCTACACATACAGCTGAAGACTTCAAAACG 1385
         || || ||||| |||||||||||||| |||  |||| |||||||||  ||||| |||||||||||
Db   842 CAACACCAATGCCGATGTACCGGTTCAACCCCCTTGACACCTCAGGACTTCAGTGG 901

Qy  1386 CATCCATGGGAGTCAACAGAGAAAAATCTGAGTCCAAGCCGTCCATATGAGACGCCCAAGTGAAATTCAT 1445
         ||||| |||||||||||||||||| ||||| |||| ||||  ||||||||||| |||||||||||| |||
Db   902 TGTCCACGGGAATCAAATGAGAAAGAGTTTCCGTTCAGAAAGTTACCAGAACCAGGTGAAGTTCAT 961

Qy  1446 CTTTGAGTTGATTCAGAATGCTGACACAGACCCAGGAGACCCAGTGTTCTCACCGTCCCACAAACT 1505
         ||||| |||||| |||||||| |||||||||||||||||||||||| |||| |||||| |||| |||
Db   962 CTTTGAGTTCATCCAAAATGCCGACACTTACAAAGAGCCAGTTCCTCATCTGCATGAACT 1021

Qy  1506 GTGA 1509
         | ||
Db  1022 ATGA 1025
```

```
RESULT 5
AC119673/c
LOCUS       AC119673              188119 bp    DNA     linear   PRI 16-JUL-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-212H11, complete sequence.
ACCESSION   AC119673
VERSION     AC119673.2  GI:21844627
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188119)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
TITLE       Direct Submission
```

JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 188119)	8696	9085	2931	3195	3145
AUTHORS	Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.	6	<800	6382	6461	2067	2183
TITLE	Direct Submission					
JOURNAL	Submitted (30-APR-2002) Genome Center, University of Washington,	2464	2523	512	<800	4754	4905
	Box 352145, Seattle, WA 98195, USA					
REFERENCE	3 (bases 1 to 188119)	14377	14333	449	<800	124	<800
AUTHORS	Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,	7675	7840	1646	1601	1936	1957
	Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.					
	and Haugen,E.D.					
TITLE	Direct Submission	4109	3986	13415	13198	3410	3414
JOURNAL	Submitted (16-JUL-2002) Genome Center, University of Washington,	1650	1641	1353	1310	9341	9345
	Box 352145, Seattle, WA 98195, USA					
COMMENT	On Jul 16, 2002 this sequence version replaced gi:20340494.	1641	1641	9257	9231	4771	4905
	--------- Genome Center					
	Center: University of Washington Genome Center					
	Center Code: UWGC	802	852	3345	3303	7995	8199
	Web site: http://www.genome.washington.edu					
	Contact: uwgc@tcgs@u.washington.edu	8094	8674	104	<800	2778	2982
	Drafting Center: SC					
	--------- Project Information					
	Center project name: chr-1	3568	3459	9998	9715	19698	19775
	Center clone name: RP11-212H11 (sc0653)					
	--------- Summary Statistics	10033	9904	2742	2741	354	<800
	Sequencing vector: plasmid; 100% of reads					
	Chemistry: Dye-terminator ET; 52% of reads	4361	4300	806	812	6508	6419
	Chemistry: Dye-terminator Big Dye; 48% of reads					
	Assembly program: Phrap; version 0.990319	259	<800	8007	8479	3029	2982
	Quality levels above 40 are expected to have less than					
	1 error in 10,000 bp.	5309	5305	1307	1310	971	955
	Base-by-base quality values are not generally visible from the					
	GenBank flat file format but are available as part	4297	4300	1065	1052	7346	7329
	of this entry's ASN.1 file.					
	---------	3090	3024	2103	2099	3818	3691
	Overlapping sequences:					
	5': RP11-131E5 (UWGC:sc0638) AL359702	6150	6153	1931	1882	10426	10302
	3': Mapping in progress	2359	2423	1631	1601	8307	8199

	Sequence Quality Assessment:	9249	9904	74	<800	5652	5672
	This entry has been annotated with sequence quality					
	estimates computed by the Phrap assembly program.	1675	1641	436	<800	2212	2183
	All manually edited bases have been reduced to quality zero.					
	Quality data were determined from quality data; all sequencing	379	<800	6462	6461	4405	4386
	problems, such as compressions and repeats; all regions were					
	covered by at least one plasmid subclone or more than one M13	592	<800	7817	7715	13198	13391
	subclone; and the assembly was confirmed by restriction digest.					
		703	<800	8986	9231	3380	3414
	Sequence Validation:					
	This sequence has been validated by Multiple Complete Digest	522	<800	7688	7715	2483	2580
	fingerprinting. Comparison of the experimentally derived digest					
	fragments with sequence-predicted fragments is given below.	1016	1018	2727	2741	3565	3414
	The electronically-digested sequence consists of both insert and					
	vector, in order to accurately represent the entire circular BAC.	1735	1641	3111	3145	14307	14103
	Small fragments below a variable cutoff (approximately 400-800 bp)					
	are not resolved in the fingerprint and hence do not appear	3979	3986	734	812	742	749
	in the table. There are no significant remaining discrepancies					
	between the experimental and predicted values. Uniquely ordered	5380	5305	4248	4170	3148	3145
	fragments are separated by dashed lines.					
		2672	2672	4484	4405	6733	6678
	EcoRI	---------	---------	---------	---------	---------	---------
		14285	14333	5369	5318	1704	1668
	HindIII	---------	---------	---------	---------	---------	---------
SeqDerMap FngrPrnt	BglII	13063	12763	6250	6461	229	<800
--------- ---------	SeqDerMap FngrPrnt					
SeqDerMap FngrPrnt	EcoRI	2586	2672	3010	3048	4457	4386
--------- ---------	--------- ---------					

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

```
LOCUS       AC079937        185081 bp   DNA   linear   HTG 19-SEP-2000
DEFINITION  Mus musculus clone RP23-135H10 strain C57BL6/J, WORKING DRAFT
            SEQUENCE, 18 unordered pieces.
ACCESSION   AC079937
VERSION     AC079937.1  GI:10190771
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 185081)
  AUTHORS   Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W.,
            Bouffard,G.G., Dietrich,N.L., Eagle,W.O., Gupta,J., Ho,S.-L.,
            Idol,J., Lee-Lin,S.-Q., Legaspi,R., Lim,M., Maduro,Q.L.,
            Maduro,V.B., Mastrian,S.D., McCloskey,J.C., McDowell,J.,
            Pearson,R., Snyder,B., Stanlripop,S., Summers,T.J., Thomas,J.W.,
            Thomas,P.J., Tlongson,E.E., Touchman,J.W., Tran,J.T., Vogt,J.L.,
            Walker,M.A., Wetherby,K.D. and Green,E.D.
  TITLE     Unpublished
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 185081)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (19-SEP-2000) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     --------------- Genome Center
            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_mouse@nhgri.nih.gov
            --------------- Project Information
            Center project name: agx
            Center clone name: 135H10
            --------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 174235 bases at least Q40
            Consensus quality: 177968 bases at least Q30
            Consensus quality: 179643 bases at least Q20
            Insert size: 218000; agarose-fp
            Insert size: 183381; sum-of-contigs
            Quality coverage: 4.04x in Q20 bases; agarose-fp
            Quality coverage: 4.80x in Q20 bases; sum-of-contigs
            ---------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1    2260: contig of 2260 bp in length
            *     2261    2360: gap of unknown length
            *     2361    5210: contig of 2850 bp in length
            *     5211    5310: gap of unknown length
            *     5311    8722: contig of 3412 bp in length
            *     8723    8822: gap of unknown length
            *     8823   11723: contig of 2901 bp in length
            *    11724   11823: gap of unknown length
            *    11824   15978: contig of 4155 bp in length
            *    15979   16078: gap of unknown length
            *    16079   18643: contig of 2565 bp in length
            *    18644   18743: gap of unknown length
            *    18744   23298: contig of 4555 bp in length
            *    23299   23398: gap of unknown length
            *    23399   30271: contig of 6873 bp in length
            *    30272   30371: gap of unknown length
            *    30372   38471: contig of 8100 bp in length
            *    38472   38571: gap of unknown length
            *    38572   47518: contig of 8947 bp in length
            *    47519   47618: gap of unknown length
            *    47619   56601: contig of 8983 bp in length
            *    56602   56701: gap of unknown length
            *    56702   68548: contig of 11847 bp in length
            *    68549   68648: gap of unknown length
            *    68649   76640: contig of 7992 bp in length
FEATURES             Location/Qualifiers
     source          1..188119
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone="RP11-212H11"
                     /clone_lib="RPCI human BAC library 11"
```

```
RESULT 6
Query Match             15.6%;  Score 235.2;  DB 9;  Length 188119;
Best Local Similarity   96.8%;  Pred. No. 3.1e-56;
Matches  240;  Conservative   0;  Mismatches   8;  Indels   0;  Gaps   0;

Qy      248 TTCATAAAAGTCTTTCCCTRCACAGTGGTCAGCCACCAGCTTTATCCAGCACATGAAAGTCGTGGAAG 307
            || | |||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||||||
Db   172804 TGCTTCTRAGTCTTTCCTRCACAGTGGTCAGCCACCAGCTTTATCCAGCCATGAAAGTCGTGGAAG 172745

Qy      308 AGTATAAGCCACCTGTTCACTATCCAAGCGTCGGAGACCCCAGCCTTGCAGCCCTACCTGCTGA 367
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172744 AGTATAGGCCACCTGTTCACTATCCAAGGCTCGGAGACCCCAGCCTTGCAGCCCTACCTGCTGA 172685

Qy      368 TGGCTCACTTTGAGTGTGTGCCCGCCCCTGAAGAAAGGCTGGGAGAGGTGCCCCCATTCTCTG 427
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172684 TGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAAGGCTGGGAGAGGTGCCCCCATTCTCTG 172625

Qy      428 GGTTGGAGCCGTGAVGGCGTCATCTATATGGTCGGGCACCACACTGGACGACAAGAACTCTGTGA 487
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   172624 GGTTGGAGCCGTGAVGGCGTCATCTATATGGTCGGGCACCACACTGGACGACAAGAACTCTGTGA 172565

Qy      488 TGGCCATTA 495
            ||| | |||
Db   172564 TGGTCTGA 172557
```

10810	---	10492	5321	---	5318	898 --- 868
397	---	<800	6530	---	6461	2021 --- 1957
476	---	<800	5041	---	5027	5046 --- 4905
4635	---	4600	9981	---	9715	1246 --- 1315
435	---	<800	2324	---	2332	1393 --- 1315
215	---	<800	1068	---	1052	
326	---	<800	4868	---	4769	
2417	---	2423	4780	---	4769	
312	---	<800	2085	---	2099	
3837	---	3783	38	---	<800	
3711	---	3783	2320	---	2332	
310	---	<800	808	---	812	
1386	---	1345	4672	---	4614	
753	---	761				
1900	---	1875				
1879	---	1875				

```
FEATURES             Location/Qualifiers
     source          1..185081
                     /organism="Mus musculus"
                     /strain="C57BL6/J"
                     /db_xref="taxon:10090"
                     /clone="RP23-135H10"
                     /clone_lib="RPCI mouse BAC library 23"
     misc_feature    1..2260
                     /note="assembly_fragment"
     misc_feature    2361..5210
                     /note="assembly_fragment"
     misc_feature    5311..8722
                     /note="assembly_fragment"
     misc_feature    8823..11723
                     /note="assembly_fragment"
     misc_feature    11824..15978
                     /note="assembly_fragment"
     misc_feature    16079..18643
                     /note="assembly_fragment"
     misc_feature    18744..23298
                     /note="assembly_fragment"
                     vector_side:right"
     misc_feature    23399..30271
                     /note="assembly_fragment"
     misc_feature    30372..38471
                     /note="assembly_fragment"
     misc_feature    38572..47518
                     /note="assembly_fragment"
     misc_feature    47619..56601
                     /note="assembly_fragment"
     misc_feature    56702..68548
                     /note="assembly_fragment"
                     clone_end:T7
     misc_feature    68649..76640
                     /note="assembly_fragment"
                     vector_side:right"
     misc_feature    76741..88724
                     /note="assembly_fragment"
     misc_feature    88825..101313
                     /note="assembly_fragment"
     misc_feature    101414..111047
                     /note="assembly_fragment"
     misc_feature    111148..136566
                     /note="assembly_fragment"
     misc_feature    136667..185081
                     /note="assembly_fragment"
                     clone_end:SP6

*     76641     76740: gap of unknown length
*     76741     88724: contig of 11984 bp in length
*     88725     88824: gap of unknown length
*     88825    101313: contig of 12489 bp in length
*    101314    101413: gap of unknown length
*    101414    111047: contig of 9634 bp in length
*    111048    111147: gap of unknown length
*    111148    136566: contig of 25419 bp in length
*    136567    136666: gap of unknown length
*    136667    185081: contig of 48415 bp in length.

BASE COUNT     48829 a   43005 c   42959 g   48571 t    1717 others
ORIGIN
```

```
Query Match     12.3%;  Score 186;  DB 2;  Length 185081;
Best Local Similarity   85.5%;  Pred. No. 5.7e-42;
Matches  207;  Conservative  0;  Mismatches  35;  Indels  0;  Gaps  0;

Qy    249  TCATAAAAGTCTTTCCTACAGTGGTTCACGACCACAGCTTATTCAGCACATGAAGTCGTGGAAGA  308
           || |||| ||| |||||||||||||| || |||| || |||||||||| ||||||||||||||
Db 162960  TCCTCAAGCCTTCCTACAGTGTTCCACAGCAGCCTTGTTCCAACATGAAGTGCGTGGCAAA  163019

Qy    309  GTATAGGCACCTGTTCACTATTCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGAT  368
           ||||||||||||||||||||| |||||||||| |||||||| |||||||| || ||||||||
Db 163020  GTATAGGCACCTGTTCACCATCCAAGGCTCAGACCCCAGTTTGCAGCCTACATGCTGAT  163079

Qy    369  GGCTCACTTGATGTGGTGCCTGCCCTGAAGAAGGCTGGGAGGGGCCCCATTCTCTCTGG  428
```

```
Db 163080  GGCTCACATTGATGTGGTGTTCCTGCCCTGAAGAAGGCTGGGAGGGGCCCCGTTCTCAGG  163139
           ||||||| ||||||||||||  ||||| |||||||||||||||||||||||| ||||| ||
Qy    429  GTTGGAGGCGTGAVGGCGTCATCTATGGTCTCGGGCGGCCACACTGGACGACGAAGAGAT  488
Db 163140  CCTGGAACGCAATTGGCTTCATCTATTGGCCACGGGGTGGTGCGCGGGTGCCTGGACAACAAAAACTTGTGAT  163199
           ||||||||| |||||||||||||  |||||||||||||||||||||||||||||| |||||
Qy    489  CCGGGAACGG  490
           ||
Db 163200  GG  163201
```

```
RESULT 7
LOCUS       AL359702    156313 bp   DNA   linear   HTG 20-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-131E5, *** SEQUENCING IN
            PROGRESS ***, 12 unordered pieces.
ACCESSION   AL359702
VERSION     AL359702.4  GI:9798027
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156313)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8977731.
            -------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            ----------------- Project Information
            Center project name: bA131E5
            ----------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 151344 bases at least Q40
            Consensus quality: 153189 bases at least Q30
            Consensus quality: 154167 bases at least Q20
            Insert size: 155213; sum-of-contigs
            Insert size: 156039; 0.2% error; agarose-fp
            Quality coverage: 4.52x in Q20 bases; sum-of-contigs Quality
            coverage: 4.72x in Q20 bases; agarose-fp
            --------------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 12 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *            1   12576: contig of 12576 bp in length
          *        12577   12676: gap of 100 bp
          *        12677   17994: contig of 5318 bp in length
          *        17995   18094: gap of 100 bp
          *        18095   54844: contig of 36750 bp in length
          *        54845   54944: gap of 100 bp
          *        54945   70607: contig of 15663 bp in length
          *        70608   70707: gap of 100 bp
          *        70708   91781: contig of 21074 bp in length
          *        91782   91881: gap of 100 bp
          *        91882  104123: contig of 12242 bp in length
          *       104124  104223: gap of 100 bp
          *       104224  106666: contig of 2443 bp in length
          *       106667  106766: gap of 100 bp
          *       106767  112670: contig of 5904 bp in length
```

```
*    112671 112770: gap of        100 bp in length
*    112771 132997: contig of 20227 bp in length
*    132998 133097: gap of        100 bp in length
*    133098 135495: contig of 2398 bp in length
*    135496 135595: gap of        100 bp in length
*    135596 153621: contig of 18026 bp in length
*    153622 153721: gap of        100 bp in length
*    153722 156313: contig of 2592 bp in length.

FEATURES             Location/Qualifiers
     source          1..156313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
                     /clone_lib="RP11-131E5"
                     /clone="RP11-131E5.1"
     misc_feature    1..12576
                     /note="assembly_fragment:01662
     misc_feature    12677..17994
                     /note="assembly_fragment:01305
     misc_feature    18095..54844
                     /note="assembly_fragment:00421
     misc_feature    54945..70607
                     /note="assembly_fragment:00007
     misc_feature    70708..9178l
                     /note="assembly_fragment:00483
     misc_feature    91882..104123
                     /note="assembly_fragment:00088
     misc_feature    104224..106666
                     /note="assembly_fragment:01682
     misc_feature    106767..112670
                     /note="assembly_fragment:00618
     misc_feature    112771..132997
                     /note="assembly_fragment:00134"
     misc_feature    133098..135495
                     /note="assembly_fragment:00993"
     misc_feature    135596..153621
                     /note="assembly_fragment:00859
     misc_feature    153722..156313
                     /note="assembly_fragment:01584
                     /vector_side:right"
                     /clone_end:T7

BASE COUNT   43117 a   35613 c  34109 g  42363 t   1111 others
ORIGIN

Query Match     11.58;  Score 173;  DB 2;  Length 156313;
Best Local Similarity  97.2%;  Pred. No. 3.3e-38;
Matches  176;  Conservative  0;  Mismatches  5;  Indels  0;  Gaps  0;

Qy  1110 CCAAGAGGTCCTTAGAACTCACGAAGAACATTGTGGCGTGATAACAGAGAGTCCAGTTCCAATGT 1169
         ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133344 CCTGGAGGTCCTPAGAACTCACGAAGAACACTGTGGCGTTGATAACAGAGAGTCCAGTTCCAATGT 133403
Qy  1170 GTTGAGTGGCCTTTGACCCCTCCCCCTCCCCCTTCTGATGACAAGGCCCTTGGGCTACCA 1229
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 133404 GTTGAGTGGCCTTTGACCCCTCCCCCGTCAGCCCTTCTCAGATGACAAGGCCCTTGGGCTACCA 133463
Qy  1230 GCTGCTCCGCGCAGACCGTACAGTCCGTCGTCTCCGGAAGTCAATATTACTGCCCCAGTTAC 1289
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||  |||
Db 133464 GCTGCTCCGCGCAGACCGTACAGTCCGTCTTCCCGGAAGTCAATATTACTGCCCCAGTAA 133523
Qy  1290 T 1290
         |
Db 133524 T 133524
```

```
gene      complement(1751..3148)
          /gene="czcS"
          /note="RSp0488; RS00366"

CDS       complement(1751..3148)
          /gene="czcS"
          /EC_number="2.7.3.-"
          /function="cell processes; protection responses"
          /note="Product confidence : probable
          Gene name confidence : probable
          predicted by Codon_usage
          predicted by Homology
          predicted by FrameD"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="PROBABLE TRANSMEMBRANE SENSORY TRANSDUCTION
          HISTIDINE KINASE FOR COBALT ZINC CADMIUM RESISTANCE
          TRANSCRIPTION REGULATOR PROTEIN"
          /protein_id="CAD17639.1"
          /db_xref="GI:17430958"
          /translation="MMWPASLAARLYLAFVLVAGVWLTGAGTYLYTALNAQLRAIDD
          AMLGKIVQLRHGCAKDTQSIDDIVRDARRYTEVSGHEGLILRVIGPDGRTIVEANPE
          GEPVAAPRAVPPDVLASESALHTWWSRGIASRAIAQARLGNHRDTVTVLVAQDGDS
          RAALMASYRRHIFLAVIAGVAAVALLSLWLVRSTLKGLARGRQAASVVPSQLDTRLT
          LEGAPRELSDLLAALNAMLTRLQDGFARLSQFSADLAHDFRPILANIIGQTEVTLAHP
          RPAGEYRAVLESSLEEYARLSRMIEDMLFLARADHARVALSVAPLDARHEIGVLVDYF
          EAVAADRDITIQVCGNGRVRADILTLLRRAINNLLDNALRHTPGGARIDVAVQVLPHST
          AIAVRNPGPGIPAASLPLLFDRFYRGDPARANSERSAGLGLAIVQTIMDLHGGTAQAS
          SPPGGLTEFRLIPPA"

gene      complement(3145..3819)
          /gene="czcR"
          /note="RSp0489; RS00367"

CDS       complement(3145..3819)
          /gene="czcR"
          /function="cell processes; protection responses"
          /note="Product confidence : probable
          Gene name confidence : probable
          predicted by Codon_usage
          predicted by Homology
          predicted by FrameD"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="PROBABLE RESPONSE REGULATOR FOR COBALT ZINC
          CADMIUM RESISTANCE TRANSCRIPTION REGULATOR PROTEIN"
          /protein_id="CAD17640.1"
          /db_xref="GI:17430959"
          /translation="MRILIVEDEPKAGDYLLKGLTESGFVADLARSGPDGLIQAMEHD
          YDLIVLDVMLPGMDGWQVIRELRRRKTTPVLFLTARDELSDRLKGLELGADDYLVKPF
          AFAELVARVHIILRRGPMRESEVLEIADLTIDTIKRRVTRAGRRIDLTAKEYALLHLL
          ARRTGEVLSRSLISSQVWDVNFDSNTNVVDVAIRRLRAKVDDPFEHKLIHTLRGMGYV
          LDTVAR"

gene      4089..4394
          /gene="RSp0490"
          /note="RS00368"

CDS       4089..4394
          /gene="RSp0490"
          /function="miscellaneous; hypothetical/global homology"
          /note="Product confidence : putative
          Gene name confidence : hypothetical
          predicted by Codon_usage
          predicted by FrameD"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="PUTATIVE SIGNAL PEPTIDE PROTEIN"
          /protein_id="CAD17641.1"
          /db_xref="GI:17430960"
          /translation="MSVRRNCLLPLAAAAFLACGAATAASVSEQALSPLQVHGARDP
          YTDGARGTRDPYTDGGAQAERATLQASVGRPDPYSDGGMSRSSAPRDRDPYTDGAQAA"

gene      4400..5725
          /gene="czcC"
          /note="RSp0491; RS00369"

CDS       4400..5725
          /gene="czcC"
          /function="cell processes; protection responses"
          /note="Product confidence : probable
          Gene name confidence : probable
          predicted by Codon_usage
          predicted by Homology
          predicted by FrameD"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="PROBABLE COBALT-ZINC-CADMIUM OUTER MEMBRANE
          RESISTANCE PROTEIN"
          /protein_id="CAD17642.1"
          /db_xref="GI:17430961"
          /translation="MVTHPCIRAAVALSCALLVRSGLAAAPAGEAAAASDTAAPAAQS
          APASVPGAAHERMTLQQAIDRALATHPALAAAGYAVRASEGAVDQASRLRNPEAAFLT
          EDVRDRATMTAQLNIPLELGSKRAARTRAAELARDAAGQDAAVLAAELRADVTRAFF
          ELSIAQERLRLAQASDALAAQEAEAAVRRVQAGKVSPVEQTKAAVARAGAAIELRDAQ
          ADLIVAQRALSAFWGGAVAPEADGSAETLPAVPPEARHRVLAAATARAEGAAIASTPRA
          LRARLEIEHRQALVNVERSKRVPDVTLSVGAKRDSGANANMAVVGIAVPLPLFDRNQG
          SLLEAQRLADKRAAEDYRALRLEQQAAALSQDVARLEAARQAVGALRQDVLPGAGRAYDA
          ARIGFEAGKFNFLDVLDAQRTLFQARARQYLAALSRAHQAQAGIDRLGR"

gene      5752..6966
          /gene="czcB"
          /note="RSp0492; RS00370"

CDS       5752..6966
          /gene="czcB"
          /function="cell processes; protection responses"
          /note="Product confidence : probable
          Gene name confidence : probable
          predicted by Codon_usage
          predicted by Homology
          predicted by FrameD"
          /codon_start=1
          /evidence=not_experimental
          /transl_table=11
          /product="PROBABLE COBALT-ZINC-CADMIUM RESISTANCE (CATION
          EFFLUX SYSTEM PROTEIN) TRANSMEMBRANE"
          /protein_id="CAD17643.1"
          /db_xref="GI:17430962"
          /translation="MTLPRQQPRATALIVALGALAGAALWWGMGARQAAHAESEAEQP
          AAHAEHAQAPAEAEARAPHAFTIAMTAEAIANAAIGVETAGAAQLRTEVVLPGEIRLN
          DDRTAHVPRVAGVAEQVAVELGQPVTRGQLLAVLSSPALSDQRSELEQAAQQRLALAQ
          RAHAREKRLWEQGIAAAQDYQQARTALQEARIADNARQKLAAIGAAPAGTALNRFEL
          RAPFDGVVVARHLSGQEAVQAEAAVFTIADLRTVWARPVFMFVSVSVESGRVPVPVA
          TATGTAVQGKVSYVGALLGEQTRSAPARVTLDNPKRAWRPGMFVSVSVESGRVPVPVA
          VRADAVQTVDEKPSVFVPVPAGFEVRPVKTGRSDGTWTEILDGLLPAGTRYAAANSYVL
          KAEAGKSADHGH"

gene      6977..10171
          /gene="czcA"
          /note="RSp0493; RS00371"

CDS       6977..10171
          /gene="czcA"
          /function="cell processes; transport of small molecules"
          /note="Product confidence : probable
          Gene name confidence : probable
          predicted by Codon_usage
```

```
Query Match    11.3%;  Score 170.4;  DB 1;  Length 213050;
Best Local Similarity   51.2%;  Pred. No. 1.9e-37;
Matches 592;  Conservative 0;  Mismatches 531;  Indels 33;  Gaps 7;

Qy  286 ATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCC 345
        |||| |||| ||  | ||| ||  |   | |  |  | ||| || ||||||||| || |||
Db  412 ATGCAGCGCGACCCGGTGGGCGATTCAGCCTGCCTCTATACGTGGAAGGCGCTCGGATCCG 471

Qy  346 AGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGTGCCTCGCCCT--------- 396
        | || ||||| |||| | |||||||||  |||  |||||||||||| || ||
Db  472 TCACTCAAGCGCGATCCTGCTGATGGCGCCACAGGACGTGGTGCCGGTGCCCCGGCACG 531

Qy  397 GAAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTGGAGCGTGGATGCGCTCATCTATGGT 456
```

```
Db    532 GAGGCGGACTGGGACCGAGCCGCCCTTTGCCGCGTCTGGGTAGAAGGACGGCATGGTGTGGGGC 591
Qy    457 CGGGGCGCACTGGACGGACAAGAAACTCTGTGATGGCCATTACTGCAGGCCCTTGGAGCTCCTG 516
           ||| |||||| |||||||||| || |||||||||| |
Db    592 CGCCGGCGCATGGGACGACAAAGGGACAACCTGATACCCCAGAATGGAAGCCCGCCGAATTGCTG 651
Qy    517 CTGATCAGGAAGTACATCCCCGAAGATCTTTTCTTCATTTTCTGGGCCATGATGAGGAG 576
           || | ||||||                      |||   | |||||||
Db    652 GCGGGCGAGCGGCGCTTCCCGTCCCCGCCCACCAATCCACTTCGGCTTCGGGTGCCGACAGGAG 711
Qy    577 TCATCAGG---GACAGGGGGCTCAGGAGGATCTCAGCGCCCTGCTACACTCAAGGGGGCGTCCAG 633
           |||||       |||||||||||||||||| ||||||||||| |||     ||||||||
Db    712 GTCGCGGCCGAGCCGGGCGGGCGGGCGAAATTGCTCGGCCTGCGGTGAAATCGCCGCGGCAGCAG 771
Qy    634 CTAGCCTTCATTGTGTGGACGAGGGGGGGCTTCATCTTGGATGATTTCATTCCTAACTTCAAG 693
           |||| |||  |  |||||  | |||||||||  |  | | ||||||
Db    772 CTCGCCTTCGTTGATCGACGAGGGGTTGCTCACCGAGGGCGGTGTGCCGGGGGCTGACC 831
Qy    694 AAGCCCATCGCCTTGATTGGAGTCTCAGAAGAAGGGTTCCATGAAACCTCATGCGCAAGTA 753
           ||| ||| ||  |  ||||||| ||| |||||||| || |||| |||| | |||||
Db    832 AAGCCGCGCGCGGCGCTGATCGGGGGCTCACCGAGAAGGGCTTCCTGTCGGTGGGCATTGAAGTTG 891
Qy    754 AACATGACTTCAGGCCACTCTTCAGCTCCTCCAAAG-----GAGACAAGCATTGGCATC 807
           |||                                      |||||||||||
Db    892 TCTGCCACGCCCCGGGCCATTCGTCGATGCCGCCCGCCCGGGGCGGGAGAGCCGGATCGCCATG 951
Qy    808 CTTGCACTGCTGTCACCCGATTGGGAGCAGAGACACCAATGCCTATCATCATATTTGGGAAGCGGG 867
           |  |                 | ||     |||||||||| | |||  ||
Db    952 ATGAGGCGCCGCTCTCAAGCACCTCGACGACGACCTGCCCATCCTGCGCCGGGTG 1011
Qy    868 ACAGTGGTGACTGTATTGCACAACTGGCAAAATGAGTTTCCCGTTCCCTGTCAATATAAATC 927
           || |||||| |  |||| | |||||||||||||||| |  |||| | ||||||
Db   1012 ACCGCCGGAGA---TGTTCGAGACGGCTGGCCCGGCGGAGATGCCGGGCTGGAACCGTGTCGCG 1068
Qy    928 CTGAGCAAACCCATGGCTATTTGAACCCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTA 987
           ||||                      | | |||  ||| |||| |||
Db   1069 CTGTGCCAACCTGTGGCGTTGTTCGAGCGCGTGCCGAAGCACTGCCAGGCGGTGCGCCCCAGC 1128
Qy    988 ACCAAATGCAATAAATCAGGACAACACTCACCATATATCCACATAATGCAGGGGTCAAGTTC 1047
           ||  ||| | |||||| | ||| |  |  | |||||||| |||||||||||||||
Db   1129 ACCAACGCCGTGCTGCAGACCACCACGCCGCTGGTCGCCGCCTGCCGGCCAACAAGGAC 1188
Qy   1048 AATGTCATCCCCCCCAGTGGGCCCAGGCCACAGTCAACTTCACCCTGGACACAGACA 1107
           ||| ||||| ||  ||||||  || |||||||||||||||||  ||||||| |||||
Db   1189 AACGTCCTGCCCGGGCCCGGAGGCAAACGGTCAACTTCCGCCGTGTCCCCGGCGACAGC 1248
Qy   1108 GTCCAAGAGGTCCTTAGAACTCACGAAGAGACATTGTGGCGTGATAAACAGAGTCCAGTTCCAT 1167
           ||||   ||       || |||||||||||| ||||| |||      || | | |||||
Db   1249 GCCTCGTCGTCCGGATCGCGGCACGTGCGGGCACGTGCGCCGGTCCCCAAGGGGCAT 1308
Qy   1168 GTGTTGAGTGCCTTTGACCCCCTC-----CCCGTCAGCCCCTTCTGATGACAAGGCCTTGG 1222
           | | |||| ||||||||||||||      |||| | |||||| ||| ||||||| |
Db   1309 TTCGAGCTGGCCCGCGCCCGTGAGCGAGGGCGCGCCTTGTCTCCGCGACGGCAATCG 1368
Qy   1223 G----CTACCAGCTGCTCTCCGGACGTACAGCTCCGTGTCTTCCCGGAAGTCAATATTACT 1278
           |     |||| | || | || ||    |||  ||||||||||  ||
Db   1369 GCGTCCTATCAGCTCATCGGCACGGTGCGCGAGGTGTTCCCG---GGGACCGTGGTG 1425
Qy   1279 GCCCCAGTTACTTCTATTGGCAAACACGAGACAGCCGATTCTTTACAAACCTCACCACTGGC 1338
           ||||||     ||  | | | ||| | | || | |  | |||  ||| ||| ||
Db   1426 GCGCCGGGGCTGATGGTGGGCGCGATTCCGGCGATCCAACCGACCAC 1485
Qy   1339 ATCTACAGGTTCTCACCCCATCTACAATACAGCCTGAAGACTTCAAACGGCATCCATGGAGTC 1398
           || | |||||||  |||||| ||||||||| | || || || |||| || | | || |||||
Db   1486 GTATTCCGGCTTCTTCTCGCCGGTGTGGCCGCCCGAGAGATCTGCGCGCCTTCACGGCACG 1545
Qy   1399 AACGAGAAAATCTCAG 1414
           |||||||||||||||
Db   1546 AACGAGGCCGCATCGG 1561
```

LOCUS   AC121225   171999 bp   DNA   linear   HTG 24-JUL-2002
DEFINITION Rattus norvegicus clone CH230-89L22, *** SEQUENCING IN PROGRESS
            ***, 77 unordered pieces.
ACCESSION   AC121225
VERSION   AC121225.3 GI:21909154
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Rattus norvegicus
  ORGANISM Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 171999)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L., Li,L.,
            Li,J., Li,Z., Lu,X., Lucier,A., Lucier,R., Liu,J., Liu,W.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mel,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,R., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,M., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,O.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 171999)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (16-MAY-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 171999)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (24-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jul 19, 2002 this sequence version replaced gi:21240469.
            -------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information

```
Center project name: GYVY
Center clone name: CH230-89L22
-----  Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye: 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 103536 bases at least Q40
Consensus quality: 116544 bases at least Q30
Consensus quality: 116544 bases at least Q30
Consensus quality: 12417 bases at least Q20
-----------------
* NOTE: Estimated insert size may differ from sequence length
*       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*       consists of 77 contigs. The true order of the pieces
*       is not known and their order in this sequence record is
*       arbitrary. Gaps between the contigs are represented as
*       runs of N, but the exact sizes of the gaps are unknown.
*       This record will be updated with the finished sequence
*       as soon as it is available and the accession number will
*       be preserved.
```

```
*      1   1095: contig of 1095 bp in length        *  40284  40383: gap of unknown length
*   1096   1195: gap of unknown length              *  40384  41805: contig of 1422 bp in length
*   1196   2367: contig of 1172 bp in length        *  41806  41905: gap of unknown length
*   2368   2467: gap of unknown length              *  41906  42923: contig of 1018 bp in length
*   2468   4128: contig of 1661 bp in length        *  42924  43023: gap of unknown length
*   4129   4228: gap of unknown length              *  43024  45164: contig of 2141 bp in length
*   4229   5398: contig of 1170 bp in length        *  45165  45264: gap of unknown length
*   5399   5498: gap of unknown length              *  45265  46850: contig of 1586 bp in length
*   5499   6503: contig of 1005 bp in length        *  46851  46950: gap of unknown length
*   6504   6603: gap of unknown length              *  46951  47962: contig of 1012 bp in length
*   6604   7990: contig of 1387 bp in length        *  47963  48062: gap of unknown length
*   7991   8090: gap of unknown length              *  48063  49548: contig of 1486 bp in length
*   8091   9719: contig of 1629 bp in length        *  49549  49648: gap of unknown length
*   9720   9819: gap of unknown length              *  49649  52368: contig of 2720 bp in length
*   9820  11874: contig of 2055 bp in length        *  52369  52468: gap of unknown length
*  11875  11974: gap of unknown length              *  52469  54148: contig of 1680 bp in length
*  11975  13052: contig of 1078 bp in length        *  54149  54248: gap of unknown length
*  13053  13152: gap of unknown length              *  54249  55775: contig of 1527 bp in length
*  13153  14280: contig of 1128 bp in length        *  55776  55875: gap of unknown length
*  14281  14380: gap of unknown length              *  55876  57739: contig of 1864 bp in length
*  14381  15777: contig of 1397 bp in length        *  57740  57839: gap of unknown length
*  15778  15877: gap of unknown length              *  57840  59807: contig of 1968 bp in length
*  15878  16879: contig of 1002 bp in length        *  59808  59907: gap of unknown length
*  16880  16979: gap of unknown length              *  59908  60918: contig of 1011 bp in length
*  16980  18537: contig of 1558 bp in length        *  60919  61018: gap of unknown length
*  18538  18637: gap of unknown length              *  61019  62525: contig of 1507 bp in length
*  18638  20054: contig of 1417 bp in length        *  62526  62625: gap of unknown length
*  20055  20154: gap of unknown length              *  62626  65179: contig of 2554 bp in length
*  20155  21580: contig of 1426 bp in length        *  65180  65279: gap of unknown length
*  21581  21680: gap of unknown length              *  65280  67322: contig of 2043 bp in length
*  21681  22789: contig of 1109 bp in length        *  67323  67422: gap of unknown length
*  22790  22889: gap of unknown length              *  67423  69197: contig of 1775 bp in length
*  22890  24429: contig of 1540 bp in length        *  69198  69297: gap of unknown length
*  24430  24529: gap of unknown length              *  69298  72316: contig of 3019 bp in length
*  24530  25932: contig of 1403 bp in length        *  72317  72416: gap of unknown length
*  25933  26032: gap of unknown length              *  72417  75706: contig of 3290 bp in length
*  26033  27792: contig of 1760 bp in length        *  75707  75806: gap of unknown length
*  27793  27892: gap of unknown length              *  75807  77956: contig of 2150 bp in length
*  27893  29087: contig of 1195 bp in length        *  77957  78056: gap of unknown length
*  29088  29187: gap of unknown length              *  78057  80390: contig of 2334 bp in length
*  29188  30744: contig of 1557 bp in length        *  80391  80490: gap of unknown length
*  30745  30844: gap of unknown length              *  80491  82296: contig of 1806 bp in length
*  30845  32736: contig of 1892 bp in length        *  82297  82396: gap of unknown length
*  32737  32836: gap of unknown length              *  82396  84424: contig of 2028 bp in length
*  32837  34268: contig of 1432 bp in length        *  84424  84524: gap of unknown length
*  34269  34368: gap of unknown length              *  84525  86700: contig of 2176 bp in length
*  34369  36154: contig of 1786 bp in length        *  86701  86800: gap of unknown length
*  36155  36254: gap of unknown length              *  86801  88030: contig of 1230 bp in length
*  36255  37586: contig of 1332 bp in length        *  88031  88130: gap of unknown length
*  37587  37686: gap of unknown length              *  88131  90375: contig of 2245 bp in length
*  37687  38996: contig of 1310 bp in length        *  90376  90475: gap of unknown length
*  38997  39096: gap of unknown length              *  90476  92063: contig of 1588 bp in length
*  39097  40283: contig of 1187 bp in length        *  92064  92163: gap of unknown length
                                                     *  92164  94087: contig of 1924 bp in length
```

```
Query Match      8.7%;      Score 131.2;  DB 2;   Length 171999;
Best Local Similarity 84.1%;  Pred. No. 4.3e-26;
Matches 148;  Conservative   0;  Mismatches   28;  Indels   0;  Gaps   0;

Qy   1115 AGGTCCTAGAACTCACGAAGAACAATTGTGGCGTGATAACAGAGTCCAGTTCCAATGTGTTGA 1174
          ||||||||||||||||||||||||||||||||||||  |||||||||||||||||||||||
Db  13680 AGGTCGTAGAAACTCGTCCAGAACACATTGTGGCGTGATGACCGAGTCCAGTTGCATGTGTTGA 13621

Qy   1175 GTGGCCTTGACCCCTCCCCGTCAGCCCTTCTGATGACCAGCCCTTGGGCGTACCAGCTGC 1234
          | |||||||||| ||  |||||||| |||||||||||||||| |||||||| ||||||
Db  13620 GATCCTTTGAACCACTGCCCGTCAGCCCCCTCTGATGACCAGCCCCATGGGCGTACCAGCTGC 13561

Qy   1235 TCCGCCAGACCGTACAGTCCGTCTTCCCGGAAGTCAATATTACTGCCCCAGTTACT 1290
          ||||| | |||| ||||||| |||| || |||| |||| |||| || |||| || |
Db  13560 TTCAACAGACCATACAGTTCGTCTTCCCGGAAGTCAAGATCAATTGTCCCGGTAAT 13505

RESULT 10
```

```
LOCUS       AL359702/c          156313 bp    DNA    linear    HTG 20-JAN-2001
DEFINITION  Homo sapiens chromosome 1 clone RP11-131E5, *** SEQUENCING IN
            PROGRESS ***, 12 unordered pieces.
ACCESSION   AL359702
VERSION     AL359702.4  GI:9798027
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      human.
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156313)
  AUTHORS   Burton,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone
            requests: clonerequest@sanger.ac.uk
COMMENT     On Aug 12, 2000 this sequence version replaced gi:8977731.
            -------------------------- Genome Center
            Center: Sanger Centre
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------------------- Project Information
            Center project name: bA131E5
            -------------------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Sequencing vector: plasmid; L08752; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Consensus quality: 151344 bases at least Q40
            Consensus quality: 153189 bases at least Q30
            Consensus quality: 154167 bases at least Q20
            Insert size: 155213; sum-of-contigs
            Insert size: 156039; 0.2% error; sum-of-contigs
            Quality coverage: 4.52x in Q20 bases; agarose-fp
            coverage: 4.72x in Q20 bases; agarose-fp
            *
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            *        1  12576: contig of 12576 bp in length
            *    12577  12676: gap of 100 bp
            *    12677  17994: contig of 5318 bp in length
            *    17995  18094: gap of 100 bp
            *    18095  54844: contig of 36750 bp in length
            *    54845  54944: gap of 100 bp
            *    54945  70607: contig of 15663 bp in length
            *    70608  70707: gap of 100 bp
            *    70708  91781: contig of 21074 bp in length
            *    91782  91881: gap of 100 bp
            *    91882 104123: contig of 12242 bp in length
            *   104124 104223: gap of 100 bp
            *   104224 106666: contig of 2443 bp in length
            *   106667 106766: gap of 100 bp
            *   106767 112670: contig of 5904 bp in length
            *   112671 112770: gap of 100 bp
            *   112771 132997: contig of 20227 bp in length
            *   132998 133097: gap of 100 bp
            *   133098 135495: contig of 2398 bp in length
            *   135496 135595: gap of 100 bp
            *   135596 153621: contig of 18026 bp in length
            *   153622 153721: gap of 100 bp
            *   153722 156313: contig of 2592 bp in length.
FEATURES             Location/Qualifiers
     source          1..156313
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="1"
```

```
     misc_feature                /clone="RP11-131E5"
                                 /clone_lib="RPCI-11.1"
     misc_feature    1..12576
                     /note="assembly_fragment:01662
                     fragment_chain:1"
     misc_feature    12677..17994
                     /note="assembly_fragment:01305
                     fragment_chain:1"
     misc_feature    18095..54844
                     /note="assembly_fragment:00421
                     fragment_chain:1"
     misc_feature    54945..70607
                     /note="assembly_fragment:00007
                     fragment_chain:1"
     misc_feature    70708..91781
                     /note="assembly_fragment:00483
                     fragment_chain:1"
     misc_feature    91882..104123
                     /note="assembly_fragment:00088
                     fragment_chain:1"
     misc_feature    104224..106666
                     /note="assembly_fragment:01682
                     fragment_chain:2"
     misc_feature    106767..112670
                     /note="assembly_fragment:00618
                     fragment_chain:2"
     misc_feature    112771..132997
                     /note="assembly_fragment:00134"
     misc_feature    133098..135495
                     /note="assembly_fragment:00993"
     misc_feature    135596..153621
                     /note="assembly_fragment:00859
                     fragment_chain:3"
     misc_feature    153722..156313
                     /note="assembly_fragment:01584
                     fragment_chain:3"
BASE COUNT    43117 a   35613 c   34109 g   42363 t    1111 others
ORIGIN
            vector_side:right"
            clone_end:t7
```

```
Query Match              8.3%;    Score 125.6;  DB 2;   Length 156313;
Best Local Similarity   97.0%;    Pred. No. 1.8e-24;
Matches 128; Conservative   0;  Mismatches   4;  Indels   0;  Gaps   0;

Qy       1378 TTCAAACCGCATCCATGGAGTCAACGAGAAAAATCTCAGTCCAAGCCTATGAGACCCAAGTG 1437
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148883 TTCCTCAGCATCCATGGAGTCAACGAGAAAAATCTCAGTCCAAGCCTATGAGACCCAAGTG 148824

Qy       1438 AAATTCATCTTTGAGTTGATTCAGAAATGCTGACACAGACCAGGAGACCCAGTTTCTCACCTG 1497
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     148823 AAAVTCATCTTTGAGTTGATTCAGAAATGCTGACACAGACCAGGAGACCCAGTTTCTCACCTG 148764

Qy       1498 CACANAACTGTGA 1509
              ||||| |||||||
Db     148763 CACANAACTGTGA 148752
```

```
RESULT 11
LOCUS       AC115001
DEFINITION  Mus musculus clone RP24-128M9, WORKING DRAFT SEQUENCE, 15 ordered
            pieces.
ACCESSION   AC115001
VERSION     AC115001.2  GI:21327577
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 134617)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-128M9
```

```
LOCUS       AC115001          134617 bp    DNA    linear    HTG 06-JUN-2002
```

JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 134617)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
           Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
           Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,J.,
           Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Kamat,A., Karatas,A., Kells,C., LaRocque,K., Lamazares,R.,
           Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
           Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,B., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (14-MAR-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 134617)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L.,
           Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
           Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,J.,
           Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
           Galagan,J., Gardyna,S., Ginde,S., Gord,S., Goyette,M., Graham,L.,
           Faro,S., Ferreira,P., FitzGerald,M., FitzHugh,W., Gage,D.,
           Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
           Grand-Pierre,N., Hagos,B., Horton,L., Hulme,W., Iliev,I.,
           Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Lindblad-Toh,K.,
           Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N.,
           Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P.,
           O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
           Pollara,V., Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C.,
           Rogov,P., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S.,
           Schupback,R., Seaman,S., Severy,B., Spencer,B., Stange-Thomann,N.,
           Stojanovic,N., Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S.,
           Theodore,J., Topham,K., Travers,M., Travis,N., Trigilio,J.,
           Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
           Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
           On Jun 6, 2002 this sequence version replaced gi:19424496.
COMMENT   All repeats were identified using RepeatMasker:
           Smit,A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           -------------------- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submissions@genome.wi.mit.edu
           -------------------- Project Information
           Center project name: L24383
           Center clone name: 128_M_9
           -------------------- Summary Statistics
           Sequencing vector: Plasmid; n/a; 100% of reads
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Assembly program: Phrap; version 0.960731
           Consensus quality: 129731 bases at least Q40
           Consensus quality: 131917 bases at least Q30
           Consensus quality: 132753 bases at least Q20
           Insert size: 134000; agarose-fp

           Insert size: 133217; sum-of-contigs
           Quality coverage: 5.6 in Q20 bases; agarose-fp
           Quality coverage: 5.6 in Q20 bases; sum-of-contigs
           * -------------------
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 15 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * provided by the submitter.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *   1   1153: contig of 1153 bp in length
           *   1154 1253: gap of 100 bp
           *   1254 3428: contig of 2175 bp in length
           *   3429 3528: gap of 100 bp
           *   3529 4365: contig of 837 bp in length
           *   4366 4465: gap of 100 bp
           *   4466 7977: contig of 3512 bp in length
           *   7978 8077: gap of 100 bp
           *   8077 12322: contig of 4245 bp in length
           * 12323 12422: gap of 100 bp
           * 12423 14768: contig of 2346 bp in length
           * 14769 14868: gap of 100 bp
           * 14869 19859: contig of 4991 bp in length
           * 19860 19959: gap of 100 bp
           * 19960 25158: contig of 5199 bp in length
           * 25159 25258: gap of 100 bp
           * 25259 33912: contig of 8654 bp in length
           * 33913 34012: gap of 100 bp
           * 34013 42066: contig of 8054 bp in length
           * 42067 42166: gap of 100 bp
           * 42166 51449: contig of 9283 bp in length
           * 51450 51549: gap of 100 bp
           * 51549 62734: contig of 11185 bp in length
           * 62735 62834: gap of 100 bp
           * 62835 80340: contig of 17506 bp in length
           * 80341 80440: gap of 100 bp
           * 80441 107825: contig of 27385 bp in length
           * 107826 107925: gap of 100 bp
           * 107926 134617: contig of 26692 bp in length.

FEATURES   Location/Qualifiers
     source   1..134617
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone="RP24-128M9"
                     /clone_lib="RPCI-24 Male Mouse BAC"
     misc_feature   1..1153
                     /note="assembly_fragment"
     misc_feature   1254..3428
                     /note="assembly_fragment"
     misc_feature   3529..4365
                     /note="assembly_fragment"
     misc_feature   4466..7977
                     /note="assembly_fragment"
     misc_feature   8078..12322
                     /note="assembly_fragment"
     misc_feature   12423..14768
                     /note="assembly_fragment"
     misc_feature   14869..19859
                     /note="assembly_fragment"
     misc_feature   19960..25158
                     /note="assembly_fragment"
     misc_feature   25259..33912
                     /note="assembly_fragment"
     misc_feature   34013..42066
                     /note="assembly_fragment"
     misc_feature   42167..51449
                     /note="assembly_fragment"
     misc_feature   51550..62734
                     /note="assembly_fragment"
     misc_feature   62835..80340
                     /note="assembly_fragment"

```
                    misc_feature    /note="assembly_fragment"
                                     80441..107825
                                     /note="assembly_fragment"
                    misc_feature    107926..134617
                                     /note="assembly_fragment"
BASE COUNT    37120 a  30277 c  28950 g  36848 t   1422 others
ORIGIN

Query Match        8.0%;    Score 120.2;  DB 2;    Length 134617;
Best Local Similarity   79.0%;  Pred. No. 6.4e-23;  DB 2;
Matches 143; Conservative   0; Mismatches 38; Indels   0; Gaps   0;

Qy   1110 CCAAGAGGTCCTAGAACTCACGAAGAACATTGTGGCTGATAACACAGAGTCCAGTTCCATGT 1169
          ||  ||||||||||||| |||||||||| ||||||  ||||||||||||||||||||||||
Db 110518 CCTGAAGGTCCTAGAACTGTCAAGAAACACCGTGGCTGCTGATGACACAGAGTCCAGTTCCATGT 110577

Qy   1170 GTTGAGTGCCTTTGACCCCCTCCCCCGTCAGCGCCTTCTGATGACACAAGGCCTTGGGGTACCA 1229
          |||||| |||||| |||||||||||||||||| |||||||| ||||||||||||||||||||||
Db 110578 GTTGAGATCCTTTGAACCCCTGCCCCATCAGCCCCTTCTGATGACCAGGCCATGGGGTACCA 110637

Qy   1230 GCTGCTCCCGCCACCGTACAGTCCGTCGTTCTCCCGGAAGTCAATATTTACTGCCCAGTTAC 1289
          ||||||||||| |||||||||||| ||||||| ||||||||||||||| || ||||||||||
Db 110638 GCTGCTTCAAGAGACCATACCGATCTGTCTTTCCCGGAAGTGGACATCGTGCTGCCCGGTAA 110697

Qy   1290 T 1290
          |
Db 110698 T 110698
```

```
RESULT 12
LOCUS       AE005748               5664 bp    DNA     linear   BCT 12-JUN-2002
DEFINITION  Caulobacter crescentus CB15 section 74 of 359 of the complete
            genome.
ACCESSION   AE005748 AE005673
VERSION     AE005748.1  GI:13421957
KEYWORDS    .
SOURCE      Caulobacter crescentus CB15.
  ORGANISM  Caulobacter crescentus CB15
            Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
            Caulobacter.
REFERENCE   1  (bases 1 to 5664)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E.,
            Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R.,
            Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D.,
            Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smith,J., Craven,M.B., Khouri,H.,
            Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and
            Fraser,C.M.
  TITLE     Complete genome sequence of Caulobacter crescentus
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
   MEDLINE   21173698
   PUBMED    11259647
REFERENCE   2  (bases 1 to 5664)
  AUTHORS   Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J.,
            Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I.,
            Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B.,
            Laub,M.T., Deboy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L.,
            Haft,D.H., Kolonay,J.F., Smith,J., Craven,M., Khouri,H., Shetty,J.,
            Berry,K., Utterback,T., Tran,K., Wolf,A., Vamathevan,J.,
            Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and
            Fraser,C.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (31-JAN-2001) The Institute for Genomic Research, 9712
            Medical Center Dr, Rockville, MD 20850, USA
FEATURES             Location/Qualifiers
     source          1..5664
                     /organism="Caulobacter crescentus CB15"
                     /strain="CB15"
                     /db_xref="taxon:190650"
     gene            complement(107..685)
                     /gene="CC0729"
```

```
     CDS             complement(107..685)
                     /gene="CC0729"
                     /note="identified by Glimmer2; putative"
                     /note="assembly_fragment"
                     /codon_start=1
                     /transl_table=11
                     /product="conserved hypothetical protein"
                     /protein_id="AAK22714.1"
                     /db_xref="GI:13421958"
                     /translation="MVILNRIYTRTGDGGTTRLATGAPVSKRASLRVDSYGYQETNNA
                     IGQARQHTRAADPVLDAILERIQNDLFDLGADLLATPEQHGKPEWEPLRVVESQVERLER
                     YLNRLSDLLFVAARWANDKGQADVLWKPGATR"
     gene            complement(723..977)
                     /gene="CC0730"
     CDS             complement(723..977)
                     /gene="CC0730"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22715.1"
                     /db_xref="GI:13421959"
                     /translation="MATVSMARGAATPYSWRMSQLFDFLVPAALLAVLIALAAGLYAL
                     FRGDFGRSWSNKLMRLRVLLQFIAVILVAAAFWMRKGA"
     gene            1281..1505
                     /gene="CC0731"
     CDS             1281..1505
                     /gene="CC0731"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22716.1"
                     /db_xref="GI:13421960"
                     /translation="MSARLATAAALSASDREEAMAEVRQLITDLRDAAQDGRQLARRY
                     ADVLQNVVDEYCVEFEKRAEAALARLEAAA"
     gene            1502..2149
                     /gene="CC0732"
     CDS             1502..2149
                     /gene="CC0732"
                     /note="identified by match to protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, TetR family"
                     /protein_id="AAK22717.1"
                     /db_xref="GI:13421961"
                     /translation="MSFVSVNIEQGAETPVKLTRRALAKQRTRERVLAAARRLFSERG
                     YEGATIRDIAQAAGMSTGAVFASPADKSELFEEIIIADYEVIYAQMIAQAARTKTVDE
                     ALLGLFGVAXSFHVEQLPLLRAS1SVSWTRSEAAERRANDLKHIFKLIGVALQRAVD
                     EGQLKKDIDAKLLAEIVWDYYVANRRAVIDGWSVEALLARLSDQLRVIFAGARA"
     gene            2233..2841
                     /gene="CC0733"
     CDS             2233..2841
                     /gene="CC0733"
                     /note="identified by match to protein family HMM"
                     /codon_start=1
                     /transl_table=11
                     /product="transcriptional regulator, TetR family"
                     /protein_id="AAK22718.1"
                     /db_xref="GI:13421962"
                     /translation="MGVRDEQKAATREKVLVAARDLFNDVGYETTIRAIAERAGVSV
                     GSVFTFASRAEVLSHVMTHRLDELYAEFDRVIPVLRGSTTDRLCSIFAIHYAFEETRR
                     LKAYAWNYRMAAAGDLPAAEMSAAMDRQIAVIAEGWKPR"
     gene            complement(3026..3148)
                     /gene="CC0734"
     CDS             complement(3026..3148)
                     /gene="CC0734"
                     /note="identified by Glimmer2; putative"
                     /codon_start=1
                     /transl_table=11
                     /product="hypothetical protein"
                     /protein_id="AAK22719.1"
```

```
gene
            /db_xref="GI:13421963"
            /translation="MMRHRATAPLGQAPEGSIWIPAFPAGKVGFLEGGCWRLFTTLLG"
            complement(3526..3744)
            /gene="CC0735"
CDS         complement(3526..3744)
            /gene="CC0735"
            /note="identified by Glimmer2; putative"
            /codon_start=1
            /transl_table=11
            /product="hypothetical protein"
            /protein_id="AAK22720.1"
            /db_xref="GI:13421964"
            /translation="MIQTTSETWCARMQAKIMAAIDAAWALIEASDDPAVLRKARDRV
            KACGELAANVRKIAAMSGQRKPAPRTPG"
gene        4026..5441
            /gene="CC0736"
CDS         4026..5441
            /gene="CC0736"
            /note="identified by match to protein family HMM"
            /codon_start=1
            /transl_table=11
            /product="M20/M25/M40 family peptidase"
            /protein_id="AAK22721.1"
            /db_xref="GI:13421965"
            /translation="MAVRTATFEAPAQADPAAARLAPARPVDVAKAAEHLAEAIRFQT
            ISHQNRADDQPAEWDKLHAWLQTTYPQAHKAMTREVVAGHGLVTTWTGSNPALAPIVL
            PQDATAWINYRIAPGDSSDKVMAKAEAVGDLPVELAFEGHRNEPSAVSSTSDAWKI
            LAGLAADESQAPVVPGLVTRAGTDSRYMGGVSSDVYRFQPLVLTVDGTKVIHGTDEHIS
            IGVAEKGYATLKVTAPAVGGHSSAPPKDGGGVVTLSKAVQAIHDNPFPMKFQGPGADM
            LKAISPHASPAVKVKPAANTWLFSSLLVKYVTAKSPAGAAAMLHTTIAPTMLKGSPKENVL
            MAHQDVVPVTPGSEGQWTHPPFAGVVADGKVWGRGAIDDKGSLVTIFFEALESVAAGGE
            KPVRIVIIVSGHDEEVRGESGAQAAAALLKSRNIKAQFVLDEGMAVVADHPVINEPAAI
            LDNVERMVRFYQRIVETAGSR"

BASE COUNT      946 a    1929 c    1888 g     901 t
ORIGIN

Query Match    7.3%;
Best Local Similarity   47.3%;
Matches   408;  Conservative    0;  Mismatches  443;  Indels   12;  Gaps   2;
           Score 110.2;  DB 1;  Length 5664;  Pred. No. 3.2e-20;
```

```
Qy    295 GAAGTCGTCGAAGAGTATAGCCACCACTGTTCACTATCCAAGGCTCGGACCCCCAGCTTGCAG 354
              |||||||||  |  |||||| |||| |||| ||||||  |||||| ||  |||  ||| ||
Db   4260 GAGGTCGTCGGCCGGCCATGGCCTCGTCTACACCTGGACCGGCTCGAACCCCGGCCCTGGCG 4319

Qy    355 CCCTACCTTGCTGATGGCCTCACTTTGATGTGGTGGCCTGCCCCCTGAAGAAGGCTGGGAGGTG 414
              ||| | || |||||||||  |||   || |||||| || || |||   |||||  |||  |||
Db   4320 CCCATCGTGCTGATGGCCACCAGAGACGTGGTGCGCCGGGTCACGCCCGGCCACGCGAGGCCAG 4379

Qy    415 ----------CCCCAATTCTCTGGGTTGGAGACGGTGAAGGGCGTCATCTATGGTCGGGGCACA 465
                      |  | ||| | |  ||| || || |||  ||||| ||  | | ||  ||||| |
Db   4380 TGGACCGACCAAGAACTCTTGCGGCCGCCGTTCGCGCGTCGCGACGGCAAAGTCTGGGCGGGCG 4439

Qy    466 CTGGACGACAAGAACTCTGTTGATGGCATTACTGCAGGCCTCTTGGAGCTCCTGCTGATCAGG 525
              ||| ||||||||| ||| |||||||||   ||| ||||| || |||||||  |||| | |||
Db   4440 ATCGACGACAAGGGCTCGCTGGTCACCATCTTCGAGGCGTTGCTGGAGGGCTTCGCG 4499

Qy    526 AAGTACATCCCCCGAAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTCATCAGG 585
              || ||| ||||| |  ||  |||| | ||| | |||||||| ||||| ||| || |||
Db   4500 GGATTCAAGCCGGGTCGCGACGCCGGCCATGAGCGCCATCGGC 4559

Qy    586 ACAGGCGGCTCAGAGATCTCAGCCCTGCTACAGTCAAAGGGCGCCTAGCCTTCATT 645
              ||||| |||||||| ||||||||| || || || |||||||||  || ||||| |
Db   4560 GAGGGCGCCAGGGCGTCGGTCATCGCCGCCGTCCCAGTCGCTTCGTG 4619

Qy    646 GTGGCAGAGGGGGCTTCAATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGCC 705
              ||||| |||| || || || | ||| || ||| ||    ||  |||||||||| |||
Db   4620 CTGGACGA---GGGCATGGCGGTGGTCGCCGACCATCCCGTCACAAATGAGCCGGCCGC 4676

Qy    706 TTGATTGCAGTCTCAGAGAAGGGGTTCCATGAAACCTACCTTGCAAGTAAACATGACTTCA 765
              || || ||  ||||||||||||| || ||||| ||||| ||||| ||    |||| ||
Db   4677 ATTATCGGCGGCTGGCCCGAGAAGGGCTATGCGACCCTGAAGTCACCGCCCCGGCGGCTG 4736

Qy    766 GGCCACTCTTCAGCTCCTCCAAAGAGGACAAGACATTGGCATCCTTGCAGCTGCTGTCAGC 825
              ||||||||  |||| || ||| |||| ||||| |  ||||| | || |   |  ||| ||
Db   4737 GGCCACTCCTCGGCCCGCGCCCGCGCCCCAAGGACGACGGCGGGCGGCGGTCGTGACCCTGTCACCCTGTCCAAGGCGGTG 4796

Qy    826 CGATTGGAGCAGACAACCACCAAATGCCTATCATATTTGGAAGCGGCGGACAGTGTTGGTGTGACTGTATTG 885
Db   4797 CAGGCCATCACGACAATCCCTTCCGATGAAGTTCCGATGAAGTTCCAGGGAGCCGGGCGGGGCGGACATGCTG 4856

Qy    886 CAGCAACTGGCAAATGAGTTTCCCGTTCCCGTCCATTATAATCCTGAGCAACCATGGGCTA 945
Db   4857 AAGGCGGATTCGGCCCCACGCGTCCACCGGTCTCCAGCAGGGTGTCTTTTGGGCCACAGCTG 4916

Qy    946 TTTGAACCACTTATAAGCAGTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGG 1005
Db   4917 TTCTCGTCGCTGCTGGTCAAGTACGGCACGGCCAAGAGCCCGGCGGGCGCCCAGTG 4976

Qy   1006 ACCACCACGGCGACTCACCATATTCAAAGCACAGGGGTCAAGTTCACAAGGTCGACCGCCCCCCAGTG 1065
Db   4977 ACCACCATCGGCGCCGACCATGCTGAAGGGCTCGCAGGAAAACGGTCCTGCCGCCGGACGG 5036

Qy   1066 GCCCAGGCCACAGTCAACTTCCCGGATTCACCCTGGACAACAGGTCCAAGAGGTCCTAGAA 1125
Db   5037 GCCACCGCCCTGGATCAACTACCGCACCGCCATCGCCCGCCCGCCGCACAGCTCCGACAAGGTGATGGCC 5096

Qy   1126 CTCACGCAAGAACATTGTTGGCGTGA 1148
Db   5097 AAGGCCAAGGAAGCGGTCGCGGCGA 5119
```

RESULT 13
LOCUS   AE008919   86564 bp   DNA   linear   BCT 11-FEB-2002
DEFINITION Uncultured proteobacterium BAC EBAC000-65D09 complete sequence.
ACCESSION   AE008919
VERSION   AE008919.1 GI:18645044
KEYWORDS   .
SOURCE   uncultured proteobacterium.
  ORGANISM uncultured proteobacterium
            Bacteria; Proteobacteria; environmental samples.
REFERENCE   1 (bases 1 to 86564)
  AUTHORS   Beja,O., Suzuki,M.T., Eisen,J.A., Fraser,C.M. and DeLong,E.F.,
            Heidelberg,J.F., Nelson,W.C., Preston,C.M.,
            Hamada,T.,
  TITLE   Unsuspected diversity among marine aerobic anoxygenic phototrophs
  JOURNAL   Nature 415 (6872), 630-633 (2002)
  MEDLINE   21822632
   PUBMED   11832943
REFERENCE   2 (bases 1 to 86564)
  AUTHORS   Beja,O., Suzuki,M.T., Heidelberg,J.F., Nelson,W.C., Preston,C.M.,
            Hamada,T., Eisen,J.A., Fraser,C.M. and DeLong,E.F.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-NOV-2001) The Institute for Genomic Research, 9712
            Medical Center Dr., Rockville, MD 20850, USA

```
FEATURES             Location/Qualifiers
     source          1..86564
                     /organism="uncultured proteobacterium"
                     /db_xref="taxon:153809"
                     /clone="BAC EBAC000-65D09"
     gene            1501..2640
                     /gene="crtB"
     CDS             1501..2640
                     /gene="crtB"
                     /note="identified by match to PFAM protein family HMM
                     PF00494"
                     /codon_start=1
                     /transl_table=11
                     /product="phytoene synthase"
                     /protein_id="AAL76346.1"
                     /db_xref="GI:18645045"
                     /translation="MSSPPKYSTKSCLMPLLSELAELPTPHTDLAHCRQTLSGGSRS
                     FWVASQLLPPTLRNDACGLYAFCRRDDLIDEGDDAAALAQLHRDLGIYSGAPQER
                     TVDRVLQRIVIHHQLPRTLLEALLEGFAWDASGRHYHTLLSDVFAYGARVAGVGVMMA
                     VLMGVRESNALARAADLGVAMQLTNIARDVGEDARAGRILPRQMLVDAGIDPESFLQ
                     APTFSPALCTVVERLLLEAEQLYRSESGIALLPVGARPGIYAARLLYAEIGHALLKS"
```

```
                  GGNSIDTRAYIGMAGKARLVLRTPQWVALNKAGLHESALPECQYLLDAVQQATAQST
                  ELHHNSVFRIYRRMVWTLDLFAVLEARQQAAASGHNNALSEGHS"

gene    2644..2805
                  /gene="EBAC000-65D09.2"
                  /note="identified by Glimmer2; putative"

CDS     2644..2805
                  /gene="EBAC000-65D09.2"
                  /note="identified by Glimmer2; putative"
                  /codon_start=1
                  /transl_table=11
                  /product="hypothetical protein"
                  /protein_id="AAL76347.1"
                  /db_xref="GI:18645046"
                  /translation="MQNRAHNSPANPIIRRLAPGLLRTVDRAIFDVSNPKTIDILIQ
                  PRILLLGSH"

gene    complement(2791..3786)
                  /gene="crtC"
                  /note="EBAC000-65D09.3"

CDS     complement(2791..3786)
                  /gene="crtC"
                  /note="EBAC000-65D09.3"
                  /note="similar to PID:1809122; identified by sequence
                  similarity; putative"
                  /codon_start=1
                  /transl_table=11
                  /product="hydroxyneurosporene synthase"
                  /protein_id="AAL76348.1"
                  /db_xref="GI:18645047"
                  /translation="MGSMPKNGPQEEIAGKALGFDAEVPDGGYRWWYLLDAFSDDGKAA
                  LTIIVFVGSVFPYIVRARRRHAGNPENFCSVRAIIVGPDRRRWSLTERGAGDLKRGC
                  DQIVIGPSRVEQSGDERTFHIHEVCNPWPTRLSGQIELSVPALGKRSFELDPDGLHR
                  WWPAAPAGRVKTIAWEGAAYLDCNAGVVVPLEETFLNWHWQRSAAEAGSGHIH
                  LDATLTNGERRSLGLRYPDMSCYELADSGRDPISSTPLWRAQRLCRTVSGQPRVIE
                  TLEESPFYARSVLDLDTGLGPRKVWHESLHLDRRTRPWMQWLLPFRMRRPLVRP"

gene    3820..4761
                  /note="EBAC000-65D09.4"

CDS     3820..4761
                  /gene="ispA"
                  /note="EBAC000-65D09.4"
                  /note="identified by match to PFAM protein family HMM
                  PF00348"
                  /codon_start=1
                  /transl_table=11
                  /product="geranyltranstransferase"
                  /protein_id="AAL76349.1"
                  /db_xref="GI:18645048"
                  /translation="MNILVLFLQQTFPSMEAYAMGKESLDCEHTLSLAIREATGQG
                  CPPILAKALDYAYPPGGARVRPRLCLAVALANGNSEPRLASAAAAAIEILHCASLVHD
                  DLPCFDDAIERRGKPSLHAAFGQRIAVLSGDALIVAAFQQLAELGSSVTHPVRMAQVT
                  AIVAAGGPMGICAGQAWESREVELEHYHQRKTGALFVAATCAGAAAAGVPSEPWR
                  ELGNSIGEAYQVADDIQDAIADTAMIGKPTGIDVALDRPSAVRELGLCQGAVKRLERCI
                  EKGLDSIPACKGRALLHDVVQQQSNRFFPQGIKHTAA"

gene    4765..5916
                  /gene="crtF"
                  /note="EBAC000-65D09.5"

CDS     4765..5916
                  /gene="crtF"
                  /note="identified by match to PFAM protein family HMM
                  PF00891"
                  /codon_start=1
                  /transl_table=11
                  /product="hydroxyneurosporene methyltransferase"
                  /protein_id="AAL76350.1"
                  /db_xref="GI:18645049"
                  /translation="MSVQQGLVSNDTPGYSTSWRLRWHQYRNKLIANPRFQKWAARTP
                  LFRLIANRRATQLHHITAGPVYSQTLAAFVELDIPKLLADGPAQTDRLTSATGIPEPA
                  RORKTALGNLWNAYGDVKRGDISPETYSELMATSQGMATQGIAAMLANFGIAAMVRHQHLYADLSDPTALL
                  GGTGAFARRAVARFPNLRATVYDLPAVTEQAAAAHINSNDTQSNALQYQGGDMFEDDLPE
                  QADIMSLVRVLHDHDKPAQHLINKAFQALSLNGQLMIAEPMAETPGSESIGHTYFGF
                  YLWAMGSGRPRTSAELTSMKKIAGPNNVREAKSMPSLVRVLIGDKINA"

gene    6037..6984
                  /gene="bchC"
                  /note="EBAC000-65D09.6"

CDS     6037..6984
```

```
                  /gene="bchC"
                  /note="identified by match to PFAM protein family HMM
                  PF00107"
                  /codon_start=1
                  /transl_table=11
                  /product="2-desacetyl-2-hydroxyethyl
                  bacteriochlorophyll-ide A dehydrogenase"
                  /protein_id="AAL76351.1"
                  /db_xref="GI:18645050"
                  /translation="MTQVATAVVFESPGNLSLREVGLPDCDAADCLVEIDYSGISTGT
                  ERLLWDGRMPHFPGLNYPLLVPGYESVSGVGRVIDACSESKIRAGSVVFVPGSRGFEDVAGL
                  FGGAARKLVVDSARLVQILQPDCGSEGVLLALTATAVHAIRKRFENGALPQLLVGHGVLG
                  RLVARICQSLGAEDLRVWEQDPARRDSGTTPVVLDKPDDTASRYERICDVSGDASILE
                  TLLPSLSKGGWIVLAGFYHKPIQDFDFIAFMSEATLRIFAAEWQPEDLEFARSLLTDNR
                  IEVQDLITHQYTAQDAAPAYDKAFSDPLCLKMILDWRAV"

gene    6981..7976
                  /gene="bchX"
                  /note="EBAC000-65D09.7"

CDS     6981..7976
                  /gene="bchX"
                  /note="identified by match to PFAM protein family HMM
                  PF00142"
                  /codon_start=1
                  /transl_table=11
                  /product="chlorophyllide reductase subunit"
                  /protein_id="AAL76352.1"
                  /db_xref="GI:18645051"
                  /translation="MSPSEAQVIDPGLEASGERADMIASSTAEAPKIQIIAIYGKGGI
                  GKSFLSNLSYMAAQQGKKVLLIQCCDPKSDTTSLLFGGKRACPTIIETAAAKKAAGDEV
                  RVEDVCKRDGVYAMELGGPEVGRDQICGGRGIIHGFETLEKLGFHDWDPDYVLLDFLGD
                  VVCGGFGLPIARDMCQKVIVVASNDLQSLYVANNVCSAVEYFRKLGGNWGVAGMVTNK
                  DDGSGEAQAPCEAVGIPELASIPANDDIRRKSANYEIIGTPESEWGSLFAGLASNVAE
                  APPAQPAPLTQDGLLELFDGDTVGRNVVLQPATLADLCDVTKLEQPSLEVVVDNV"

gene    7951..9468
                  /gene="bchY"
                  /note="EBAC000-65D09.8"

CDS     7951..9468
                  /gene="bchY"
                  /note="identified by match to PFAM protein family HMM
                  PF00148"
                  /codon_start=1
                  /transl_table=11
                  /product="chlorophyllide reductase subunit"
                  /protein_id="AAL76353.1"
                  /db_xref="GI:18645052"
                  /translation="MRSSTTMSDHASNSPETIPLVAEDATDAPGCAGSSRTLRDQAAG
                  QANDETLDQYAKDYPVGPHDPQQSMCPAPGSLNGVLRQRTATVLSGSACCVGLTFT
                  SHFVGARRTVGYVPFDSESLVTGKLFEDIRESVHTLADPAKYDAPAKVIAVVINLCYTAYAGSGV
                  PIDLLPKSINGVRIIGIDVPGFGVPTHAEAKDVLAAAMLKYARSEVSAGCPVARPTSLE
                  SRDSKPSIALLGEIFPVDAIVGRDLLRGLGAELAGVAVPTPREWRELYAALDSSIVAMLH
                  PFYSACNREFVAAGKPVIGSAPVGAEGTAREWLTAIGTALGTPADVWTKHRDLIRKSIA
                  EAIAANPIKARMTLSGYEGSELLVGRLLVESGADLRYVGSSACPGTKWNAHDKEWLEAH
                  GVTVKFRASLEDDTCAVDAFEPDLAIGTTPVVDKAHKQKGVPSLYFTNIISARDLMGPA
                  GAGSLSQVINTAVNNKSRFAAMDSFFEGVGSGDAAGVWTEDTINRSGLIKTREVA"

gene    9468..10910
                  /gene="bchZ"
                  /note="EBAC000-65D09.9"

CDS     9468..10910
                  /gene="bchZ"
                  /note="identified by match to PFAM protein family HMM
                  PF00148"
                  /codon_start=1
                  /transl_table=11
                  /product="chlorophyllide reductase subunit z"
                  /protein_id="AAL76354.1"
                  /db_xref="GI:18645053"
                  /translation="MLVFDEHDRAGGYWGSVYVFTAVKGLQVIIDGPVGGCENLPVTSVL
                  HYTDGLPPHELPIVVTGLAEEELGQKTGEGALAHEQTLDAERPSAVVTGSIAEMIGG
                  GVTPADTNIKRFLPRTIDEDQWQSADRALVWLMDQYGGEKKKRKRAAKENAVPLVNIIG
                  PAYGFYWTASDQAEIVRLVEGMGARVNRVPFLGCHLDDIPSLGPDADANVCMTREFGSK
                  LCQALEKPWPGAPFGLHSTTTKFLRALGEELGDPEPFIEQERHTTLKPVWDLMRSVTQ
                  DFFGTANFGITAATETYTRGVRHFLEDELGLPCAEHFESRKPGVKFDNETIRQTIAETSP
                  LILFGSYNERMYLAELGGKSIYIPASLPGTIIRRHTGPFMGIYSGATVILLQEVCNALF
                  DALFHILPLGTEMDAVEAFPARMALRWADDAEKILETLVAKQPLLVRISAAKIRDAA
                  EHSARHQGVDAVSKEHVLRAGLEVKTGTAA"
```

```
gene    11242..12066
        /gene="pufL"
        /note="EBAC000-65D09.10"
CDS     11242..12066
        /gene="pufL"
        /note="identified by match to PFAM protein family HMM
        PF00124"
        /codon_start=1
        /transl_table=11
        /product="photosynthetic reaction center L subunit"
        /protein_id="AAL76355.1"

Query Match   7.0%;  Score 106.2;  DB 1;  Length 85564;
Best Local Similarity   47.2%;
Matches  392;  Conservative  0;  Mismatches  433;  Indels  6;  Gaps  2;

Qy    395 CTGAAGAAGGCTGGGAGGTGCCCCAATTCTCTGGGTTGGACCGTGATGGGCGTCATCTATG 454
          | |||| |||||  | ||  ||||| | ||| | |||||| |  ||  ||  ||  |  | |
Db  61150 CAGAAGACGGCCTGGCAGCACCCGCCATTTGCTGGGCGTCATCGAAAATAATAATCTATACG 61209

Qy    455 GTCGGGGCACACTGAGCACGACAAGAACTCTGTGATGGCATTACTGACGGCCTTGGAGCTCC 514
          |||||  | ||  |  | ||  || || | | |   || |  | | | |  || |  | ||
Db  61210 GTCGCGGCACTTTAGATGACAAGCAAGGCCGGTTCTCTGAGGCCTTCTAGAAGCCACCGAAACTC 61269

Qy    515 TGCTGATCAGGAAGTACATCCCCCGAAGAGATCTTTCTTCATTTCTCTGGGGCCATGATGAGG 574
          ||||  ||||||||  | ||| || ||  |  |  | ||  | |  | |||| ||  |  | |
Db  61270 TTCTGAAAGAGGTTATCAGCCCCCCAGGACTTAGTCTTCTGGCGTTGGTCACGACGAGG 61329

Qy    575 A---GTCATCAGGACAGCAGGGGCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGCGTCC 631
              |||  |||| |||| |  ||| | ||| || | | |||  | | |||| ||  |  |||
Db  61330 AAATCGGAGTGGGAAATGGCCGCAGCCCCGAGCCCAGCCATGAGGGAACTTAATCTCT 61389

Qy    632 AGCTAGCCTTCATTGTGGGACGAGGGGGGGGGTTCATCTTTGGATGATTTCATTCCTAACTTCA 691
          ||||  || | | || ||| ||  || ||||| | |||  || |  |  | |  ||  |  ||
Db  61390 CCTTTGACTGGATTGGTCGACGAAGGGCGGTTTGTTGTTGTACGGCGACACCCCTTTGCTGCTG 61449

Qy    692 AGAAGCCCCATCGCCTTGATTGCAGTCTCACGAAGAGGGTTCCATGAAGCTCATGCTGTGCAAG 751
          |||  |||| || |  || ||  | | | | |||| || || |  || || ||| |  | |||
Db  61450 GACGAGACCTCGCAATGATTAATGTGGGCCGAAAAGGGTTATGTCACGCTGCACACTTACCA 61509

Qy    752 TAAACATGACTTCAGGCACCACTCTTCAGCTCCTCCAAAGGAGGACAAGCATTGGCATCCTTG 811
          | ||| | || |  ||  |  | | ||  | |||  |  | |||||| || ||  |   ||
Db  61510 CACAGGCACCGGAGGTCACTCATCGAGCCCAGAGCACAGGCCCCATAGGAACCCTCG 61569

Qy    812 CAGCTGCTGTGCAGCCGATTGGAGCAGACACCAAATCCTATCATTATTTGAAGCGGGACAG 871
          || ||| | | |  | || ||| |||| |   ||  ||  | |  | |  ||  |||| |
Db  61570 CTAGGGCACTTGATCGCATCGGAGAGACCCTTCGACGGCCCGAAGCTAGT---TGAGCCTA 61626

Qy    872 TGGTGACTGTATTGCAGCAACTGGGGCAAAATGAGTTTCCCTTCCCTGTCAAATATAATCCTGA 931
          | ||| | | | |  || || | ||| |  | | ||| | | || |  |   || |   | |||
Db  61627 TGCGCAGTGCTCTCACTATGATGGTCGCCAGAAATGGCTCCCAGAGCGATTTCTCATTCA 61686

Qy    932 GCAACCCATCGCTATTTGAACACCTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCA 991
          ||  ||| ||| |  | || | |  | |  |||| |  | |||| | |   || | |  |
Db  61687 ACAACTATGCCTTTTTCGACAGCCTCATCCTGCTGGGCAAAATGGGACATACCACCACCC 61746

Qy    992 ATGCAATAATCAGGACCACCACGGGCCGACCACATTCAAAGCAGGGGGTCAAGTTCAATG 1051
          | |||| |  ||||| || |  ||||| || |   | ||| | |||  ||||||  |  |
Db  61747 AGCCCATGGTAAGGACAACACACCACGGGCCGACCACATGTTAACGCGGGGTCAAGTTCAATG 61806

Qy   1052 TCATCCCCCAGTGCCCAGCCCAGTCAACTTCCCGGATTCACCCTGGACAGCAGAGTCC 1111
          ||||  ||  | | | || | | | |||| ||  ||  | |  | ||  | ||| ||
Db  61807 TGATACCCACCGGTGCTGAAGCCCAAGGTCAAATTTTCGACTGTGCTTACCCGGGGACACTGTAG 61866

Qy   1112 AAGAGGTCCTTAGAACTCACGAAGAACATTGTGGCTGTGATTAACGACGAGTCCAGTTCCATGTGT 1171
          | || ||| || |  |||  |  |  ||  | |  || |  |  |  | |  ||| || ||  | |
Db  61867 ACATGTTTGATCGCCACGATCACCGAAAATTGTCGACGATCACAAAGGTTGGTTTGTACTCTACG 61926

Qy   1172 TGAGTGCCTTTGACCCCCCGGTTCGGGTGTTGGGGGGGATGGATGCCCGGTGGTGGGGGGTTG 1222
          |  |  |||||||||  | |  |   |   |  |  ||| | ||| | |  | |  ||||  ||
Db  61927 ATAGGTGGATGGATCGTCCCTGGGTGTTGGGGGATGCCAAGGCAACGGCTTTG 61977
```

```
        Contact: hgsc-help@bcm.tmc.edu
---------------------
                  Project Information
Center project name: GYVY
Center clone name: CH230-89L22
---------------------
                  Summary Statistics
Sequencing vector: Plasmid;
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly Program: Phrap; version 0.990329
Consensus quality: 103536 bases at least Q40
Consensus quality: 116344 bases at least Q30
Consensus quality: 124217 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*  consists of 77 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.

*      1   1095: contig of 1095 bp in length
*   1096   1195: gap of unknown length
*   1196   2367: contig of 1172 bp in length
*   2368   2467: gap of unknown length
*   2468   4128: contig of 1661 bp in length
*   4129   4228: gap of unknown length
*   4229   5398: contig of 1170 bp in length
*   5399   5498: gap of unknown length
*   5499   6503: contig of 1005 bp in length
*   6504   6603: gap of unknown length
*   6604   7990: contig of 1387 bp in length
*   7991   8090: gap of unknown length
*   8091   9719: contig of 1629 bp in length
*   9720   9819: gap of unknown length
*   9820  11874: contig of 2055 bp in length
*  11875  11974: gap of unknown length
*  11975  13052: contig of 1078 bp in length
*  13053  13152: gap of unknown length
*  13153  14280: contig of 1128 bp in length
*  14281  14380: gap of unknown length
*  14381  15777: contig of 1397 bp in length
*  15778  15877: gap of unknown length
*  15878  16879: contig of 1002 bp in length
*  16880  16979: gap of unknown length
*  16980  18537: contig of 1558 bp in length
*  18538  18637: gap of unknown length
*  18638  20054: contig of 1417 bp in length
*  20055  20154: gap of unknown length
*  20155  21580: contig of 1426 bp in length
*  21581  21680: gap of unknown length
*  21681  22789: contig of 1109 bp in length
*  22790  22889: gap of unknown length
*  22890  24429: contig of 1540 bp in length
*  24430  24529: gap of unknown length
*  24530  25932: contig of 1403 bp in length
*  25933  26032: gap of unknown length
*  26033  27792: contig of 1760 bp in length
*  27793  27892: gap of unknown length
*  27893  29087: contig of 1195 bp in length
*  29088  29187: gap of unknown length
*  29188  30744: contig of 1557 bp in length
*  30745  30844: gap of unknown length
*  30845  32736: contig of 1892 bp in length
*  32737  32836: gap of unknown length
*  32837  34268: contig of 1432 bp in length
*  34269  34368: gap of unknown length
*  34369  36154: contig of 1786 bp in length
*  36155  36254: gap of unknown length
*  36255  37586: contig of 1332 bp in length
*  37587  37686: gap of unknown length
*  37687  38996: contig of 1310 bp in length
*  38997  39096: gap of unknown length
*  39097  40283: contig of 1187 bp in length
*  40284  40383: gap of unknown length
*  40384  41805: contig of 1422 bp in length
*  41806  41905: gap of unknown length
*  41906  42923: contig of 1018 bp in length
*  42924  43023: gap of unknown length
*  43024  45164: contig of 2141 bp in length
*  45165  45264: gap of unknown length
*  45265  46850: contig of 1586 bp in length
*  46851  46950: gap of unknown length
*  46951  47962: contig of 1012 bp in length
*  47963  48062: gap of unknown length
*  48063  49548: contig of 1486 bp in length
*  49549  49648: gap of unknown length
*  49649  52368: contig of 2720 bp in length
*  52369  52468: gap of unknown length
*  52469  54148: contig of 1680 bp in length
*  54149  54248: gap of unknown length
*  54249  55775: contig of 1527 bp in length
*  55776  55875: gap of unknown length
*  55876  57739: contig of 1864 bp in length
*  57740  57839: gap of unknown length
*  57840  59807: contig of 1968 bp in length
*  59808  59907: gap of unknown length
*  59908  60918: contig of 1011 bp in length
*  60919  61018: gap of unknown length
*  61019  62525: contig of 1507 bp in length
*  62526  62625: gap of unknown length
*  62626  65179: contig of 2554 bp in length
*  65180  65279: gap of unknown length
*  65280  67322: contig of 2043 bp in length
*  67323  67422: gap of unknown length
*  67423  69197: contig of 1775 bp in length
*  69198  69297: gap of unknown length
*  69298  72316: contig of 3019 bp in length
*  72317  72416: gap of unknown length
*  72417  75706: contig of 3290 bp in length
*  75707  75806: gap of unknown length
*  75807  77956: contig of 2150 bp in length
*  77957  78056: gap of unknown length
*  78057  80390: contig of 2334 bp in length
*  80391  80490: gap of unknown length
*  80491  82296: contig of 1806 bp in length
*  82297  82396: gap of unknown length
*  82397  84424: contig of 2028 bp in length
*  84425  84524: gap of unknown length
*  84525  86700: contig of 2176 bp in length
*  86701  86800: gap of unknown length
*  86801  88030: contig of 1230 bp in length
*  88031  88130: gap of unknown length
*  88131  90375: contig of 2245 bp in length
*  90376  90475: gap of unknown length
*  90476  92063: contig of 1588 bp in length
*  92064  92163: gap of unknown length
*  92164  94087: contig of 1924 bp in length
```

```
Query Match   6.6%;  Score 100.2;  DB 2;  Length 171999;
Best Local Similarity  83.2%;
Matches  114;  Conservative  0;  Mismatches  23;  Indels  0;  Gaps  0;

Qy    693 GAAGCCCATCGCCTTGATTGCAGTCTCAGAGAAGGGTTCCATGAAACCTTCATGCTGCAAGT 752
          ||| | |||| |||| |||| | || ||||||||| |||||||||||| |||||||||||||
Db 159979 GATCCTCCCCACAGGATTTCAGTCACCGAGAAGAAGGGTGCCCTTGACCTCAATCATCCTGCAAGT 160038

Qy    753 AAACATGACTTCAGGCCACTCTTCAGCTGTCCTTCCAAAGAGACAAAGCATTGGCATCCTTGC 812
          |||||||||||||||||||||||||||||||  |||||||||||| || |||||| |||||||
Db 160039 AAACATGACTTCAGGCCACTCTTCAGCTGTCCCCCAAAGAGACAAGCAAGCATTGGAATCCTTTC 160098

Qy    813 AGCTGCTGTCAGCCGGAT 829
          | ||||||||||||||| |
Db 160099 TGCCGCTGTCAGCCGGT 160115
```

```
RESULT 15
LOCUS       AC119673      188119 bp    DNA     linear   PRI 16-JUL-2002
DEFINITION  Homo sapiens chromosome 1 clone RP11-212H11, complete sequence.
ACCESSION   AC119673 AL365208
VERSION     AC119673.2  GI:21844627
KEYWORDS    HTG.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 188119)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 188119)
AUTHORS     Kaul,R.K., Olson,M.V., Raymond,C. and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (30-APR-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 188119)
AUTHORS     Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUL-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On Jul 16, 2002 this sequence version replaced gi:20340494.
            -------------- Genome Center
            Center: University of Washington Genome Center
            Center Code: UWGC
            Web site: http://www.genome.washington.edu
            Contact: uwgchtgs@u.washington.edu
            Drafting Center: SC
            -------------- Project Information
            Center project name: chr-1
            Center clone name: RP11-212H11 (sc0653)
            -------------- Summary Statistics
            Sequencing vector: plasmid; 100% of reads
            Chemistry: Dye-terminator ET; 52% of reads
            Chemistry: Dye-terminator Big Dye; 48% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 187572 bases at least Q40
            Consensus quality: 188068 bases at least Q30
            Insert size: 188119; sum-of-contigs
            Quality coverage: 9.0x in Q20 bases; sum-of-contigs
            --------------
            Overlapping Sequences:
            5':  RP11-131E5 (UWGC:sc0638) AL359702
            3':  Mapping in progress
            --------------
            Sequence Quality Assessment:
            This entry has been annotated with sequence quality
            estimates computed by the Phrap assembly program.
            All manually edited bases have been reduced to quality zero.
            Quality levels above 40 are expected to have less than
            1 error in 10,000 bp.
            Base-by-base quality values are not generally visible from the
            GenBank flat file format but are available as part
            of this entry's ASN.1 file.
```

This sequence was finished as follows unless otherwise noted:
all regions were either double-stranded or sequenced with an
alternate chemistry or covered by high quality data (i.e., Phred
quality >= 30); an attempt was made to resolve all sequencing
problems, such as compressions and repeats; all regions were
covered by at least one plasmid subclone or more than one M13
subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
This sequence has been validated by Multiple Complete Digest
fingerprinting. Comparison of the experimentally derived digest
fragments with sequence-predicted fragments is given below.
The electronically-digested sequence consists of both insert and
vector, in order to accurately represent the entire circular BAC.
Small fragments below a variable cutoff (approximately 400-800 bp)
are not resolved in the fingerprint and hence do not appear
in the table. There are no significant remaining discrepancies
between the experimental and predicted values. Uniquely ordered
fragments are separated by dashed lines.

EcoRI		HindIII		BglII	
SeqDerMap	FngrPrint	SeqDerMap	FngrPrint	SeqDerMap	FngrPrint
8696	9085	2931	2918	3195	3145
6	<800	6382	6461	2067	2183
2464	2523	512	<800	4754	4905
14377	14333	449	<800	124	<800
7675	7840	1646	1601	1936	1957
4109	3986	13415	13198	3410	3414
1650	1641	1353	1310	9341	9345
1641	1641	9257	9231	4771	4905
802	852	3345	3303	7995	8199
8094	8674	104	<800	2778	2982
3568	3459	9998	9715	19698	19775
10033	9904	2742	2741	354	<800
4361	4300	806	812	6508	6419
259	<800	8007	8479	3029	2982
5309	5305	1307	1310	971	955
4297	4300	1065	1052	7346	7329
3090	3024	2103	2099	3818	3691
6150	6153	1931	1882	10426	10302
2359	2423	1631	1601	8307	8199
9249	9904	74	<800	5652	5672
1675	1641	436	<800	2212	2183
379	<800	6462	6461	4405	4386
592	<800	7817	7715	13198	13391
703	<800	8986	9231	3380	3414
522	<800	7688	7715	2483	2580
2993	3024	271	<800	2597	2777
4534	4468	5092	5027	5669	5672
7872	7840	2669	2741	5005	4905

```
FEATURES
    source          Location/Qualifiers
                    1..188119
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /chromosome="1"
                    /clone="RP11-212H11"
                    /clone_lib="RPCI human BAC library 11"

Query Match        6.2%;    Score 93.2;  DB 9;   Length 188119;
Best Local Similarity  96.9%;   Pred. No. 4.5e-15;
Matches   95;  Conservative   0;  Mismatches  3;   Indels  0;  Gaps  0;

Qy        159 GGCGCTGAAAGGTGCCATCCAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATAC 218
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     181015 GACTCTGACAGGTGCCATCCAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATAC 181074

Qy        219 TACAGCCCTGGCTGAGTTCGGAAATACATTCAATAAAG 256
              ||||||||||||||||||||||||||||||||||||||
Db     181075 TACAGCCCTGGCTGAGTTCGGAAAATACATTCAATAAAG 181112
```

Qy	Db			
844	852	8570	8909	5900 5924
1016	1018	2727	2741	3565 3414
1735	1641	3111	3145	14307 14103
3979	3986	734	812	742 749
5380	5305	4248	4170	3148 3145
2672	2672	4484	4405	6733 6678
14285	14333	5369	5318	1704 1668
13063	12763	6250	6461	229 <800
2586	2672	3010	3048	4457 4386
10810	10492	5321	5318	898 868
397	<800	6530	6461	2021 1957
476	<800	5041	5027	5046 4905
4635	4600	9981	9715	1246 1315
435	<800	2324	2332	1393 1315
215	<800	1068	1052	
326	<800	4868	4769	
2417	2423	4780	4769	
312	<800	2085	2099	
3837	3783	38	<800	
3711	3783	2320	2332	
310	<800	808	812	
1386	1345	4672	4614	
753	761			
1900	1875			
1879	1875			

Search completed: July 1, 2003, 07:06:46
Job time : 4054 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: July 1, 2003, 02:38:10 ; Search time 389 Seconds
                         (without alignments)
                         8735.901 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1 atggctcagcggtgcgttg..........ctcacctgcacaaactgtga 1509
Scoring table: IDENTITY_NUC
               Gapop 10.0 , Gapext 1.0
Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   N_Geneseq_101002.*

1:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*			
2:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*			
3:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*			
4:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*			
5:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*			
6:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*			
7:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*			
8:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*			
9:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*			
10:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*			
11:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*			
12:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*			
13:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*			
14:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*			
15:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*			
16:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*			
17:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*			
18:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*			
19:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*			
20:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*			
21:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*			
22:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*			
23:	/SIDS2/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*			
24:	/SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*			

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1310.2	86.8	1421	24	AAS97192	Human metalloprote
2	496.4	32.9	500	21	AAC74554	Human ORFX ORF109
3	267	17.7	319	20	AAX40258	Human secreted pro
4	251	16.6	348	24	ABN24990	Human ORFX polynuc
5	245	16.2	362	22	AAI84929	Human polynucleoti
6	51	3.4	22	22	AAI73675	Human silent SNP c
7	50.8	3.4	1221	14	AAO33105	Pig coding sequenc
8	49.4	3.3	51	22	AAI73672	Human silent SNP c
9	49.4	3.3	51	22	AAI73674	Human silent SNP c

10	49.2	3.3	1224	14	AAO33106	Human coding seque
11	49.2	3.3	1415	24	ABK83807	Human cDNA differe
12	47.8	3.2	51	22	AAI73673	Human silent SNP c
13	46.2	3.1	381	24	ABN75263	Human ORF210 cDNA,
14	46	3.0	1445	24	ABL60776	cDNA encoding an e
15	44.8	3.0	663	24	ABK30982	Plant dwarfing/stu
16	44.8	3.0	787	24	ABK30623	Plant dwarfing/stu
17	44.8	3.0	799	24	ABK30726	Plant dwarfing/stu
18	44.8	3.0	1493	21	AAC33510	Arabidopsis thalia
19	44.8	3.0	1494	21	AAC50085	Arabidopsis thalia
20	44.8	3.0	1612	24	ABK30880	Plant dwarfing/stu
21	42.2	2.8	767	21	AAC43825	Arabidopsis thalia
22	42	2.8	495	24	AAD33895	Human carboxypepti
23	42	2.8	536	24	AAD33904	Human carboxypepti
24	42	2.8	613	24	AAD33890	Human carboxypepti
25	42	2.8	623	24	AAD33886	Human carboxypepti
26	42	2.8	640	24	AAD33889	Human carboxypepti
27	42	2.8	665	24	AAD33893	Human carboxypepti
28	42	2.8	754	22	AAH06896	Human carboxypepti
29	42	2.8	1185	21	AAC77081	Human ORFX ORF2636
30	42	2.8	1428	22	AAH75199	Nucleotide sequenc
31	42	2.8	1428	24	AAD33906	Human carboxypepti
32	42	2.8	1428	24	AAS97193	Human metalloprote
33	42	2.8	2643	22	AAH14599	cDNA sequence
34	42	2.8	2643	24	AAD33907	Human carboxypepti
35	42	2.8	2710	22	AAI50038	Human polynucleot
36	41.6	2.8	1988	22	AAI99591	Human polynucleot
37	41.6	2.8	1988	24	AAS09224	Novel cDNA encodin
38	41.6	2.8	1997	21	AAC78161	Novel cancer assoc
39	40.6	2.8	558	24	AAG33892	Human carboxypeptl
40	40.4	2.7	1879	24	AAS03019	Human carboxypeptl
41	40.4	2.7	2221	22	AAH75171	Nucleotide sequenc
42	40.4	2.7	2657	22	AAI60824	Human diagnostic a
43	40.4	2.7	2659	22	AAI99566	Human polynucleotl
44	39.8	2.6	594	24	AAD33887	Human expressed po
45	39.8	2.6	14510	24	ABN59954	Human carboxypeptl

RESULT 1
AAS97192

ID   AAS97192 standard; cDNA; 1422 BP.
XX
AC   AAS97192;
XX
DT   26-FEB-2002 (first entry)
XX
DE   Human metalloprotease partial DNA sequence #21.
XX
KW   Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW   vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
KW   hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
KW   anorectic; antiinflammatory; aspartyl protease; cysteine protease;
KW   metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
KW   lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;
KW   immune-related disease; cardiovascular disease; neuronal disease;
KW   migraine; sexual dysfunction; mood disorder; attention disorder;
KW   cognition disorder; hypotension; hypertension; psychotic disorder; ss.
OS   Homo sapiens.
XX
PN   WO200183782-A2.
XX
PD   08-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-US14431.
XX
PR   04-MAY-2000; 2000US-201879P.
XX
PA   (SUGE-) SUGEN INC.

ALIGNMENTS

XX
PI   Plowman, GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI   Payne V;
XX
DR   WPI; 2002-041502/05.
DR   P-PSDB; AAU72909.
XX
PT   Novel protease polypeptide useful for screening for substances that may
PT   be used to treat, e.g., cancers, immune-related diseases,
PT   cardiovascular disease, migraine, pain, psychotic and inflammatory
PT   disorders -
XX
PS   Claim 30; Figure 1V; 232pp; English.
XX
CC   The invention relates to an isolated, enriched, or purified protease
CC   polypeptide (I) and polynucleotide (II) encoding (I). (I) may be used to
CC   screen for substances (S) that may modulate its activity. Administering
CC   S (which modulates protease activity in vitro) may be used to treat a
CC   disease or disorder selected from cancers (e.g. of tissues, of blood or
CC   haematopoietic origin, of the breast, colon, lung, prostrate, cervical,
CC   brain, ovarian, bladder or kidney), immune-related diseases and
CC   disorders, cardiovascular disease, brain or neuronal-associated diseases
CC   (e.g. central or peripheral nervous system diseases, migraine, pain,
CC   sexual dysfunction, mood disorders, attention disorders, cognition
CC   disorders, hypotension, hypertension, psychotic disorders, neurological
CC   disorders and dyskinesias), metabolic disorders and inflammatory
CC   disorders. (I) may also be useful as a diagnostic tool for a disease or
CC   disorder such as those above. AAS97159-AAS97195 represent human
CC   protease coding sequences and primers of the invention.
XX
SQ   Sequence 1422 BP; 352 A; 390 C; 350 G; 330 T; 0 other;

Query Match   86.8%;   Score 1310.2;   DB 24;   Length 1422;
Best Local Similarity   94.0%;
Matches 1419;   Conservative 0;   Mismatches 3;   Indels 87;   Gaps 2;

RESULT 2
AAC74554
AAC74554
ID   AAC74554 standard; cDNA; 500 BP.
XX
AC   AAC74554;

```
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF109 polynucleotide sequence SEQ ID NO:217.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive; ss.
XX
OS   Homo sapiens.
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999;  99US-0127607.
PR   02-APR-1999;  99US-0127636.
PR   05-APR-1999;  99US-0127728.
PR   30-MAR-2000;  2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-602362/57.
DR   P-PSDB; AAB40345.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 5; Page 546; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; antiarthritic; immunosuppressant;
CC   osteopathic; thrombolytic; coagulant; vasotropic;
CC   immunostimulant; cardiant; dermatological; immunosuppressive;
CC   antidiabetic; hypotensive; antibacterial; antiviral; antifungal; antirheumatic;
CC   antiinflammatory; antithyroid; and antianaemic. The sequences can be used for determining
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence 500 BP; 114 A; 131 C; 136 G; 117 T; 2 other;

Query Match             32.9%;  Score 496.4;  DB 21;  Length 500;
Best Local Similarity   99.8%;
Matches  497;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy    141 ACGGGTCGGCGATGAAAGAGGGCGCTGAAAAGGTGCCATCCAACAGTGACTTTTAG 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3 ACGGGTCGGCGATGAAAGAGGGCGCTGAAAAGGTGCCATCCAACAGTGACTTTTAG 62

Qy    201 CTCTGGAGGAGTCCAATACTACACAGCCCCTGGCCTGTGAGTTCGGAAAAATACATTCATAAAGTCTT 260
          |||||||| |||||||||||||  ||||||||||| |||||||||||||||||||||||||||||||
Db     63 CTCTGGAGAAGTCCAATACTACAGCGCCCTGGCTGAGTTCGGAAAAATACATTCATAAAGTCTT 122

Qy    261 TCCTACAGTGGTCAGCACCAGCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCT 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    123 TCCTACAGTGGTCAGCACCAGCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCT 182

Qy    321 GTTCACTATCCAAGGCGTCGGACCCCAGCCTTGCAGCCCTACCTGCTGATGGCTCACTTTGA 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 GTTCACTATCCAAGGCGTCGGACCCCAGCCTTGCAGCCCTACCTGCTGATGGCTCACTTTGA 242

Qy    381 TGTGGTGCCGTGCCCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTCGTCGGTTGAGCGTGA 440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    243 TGTGGTGCCGTGCCCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTCGTCGGTTGAGCGTGA 302

Qy    441 TGGCGTCATCTATGGTTGGGGGCACACTGGACACAAGAGAACTCTGTTGATGGCATTACTGCA 500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    303 TGGCGTCATCTATGGTTGGGGGCACACTGGACACAAGAGAACTCTGTTGATGGCATTACTGCA 362

Qy    501 GGCCTTGGAGCTCCTCGTCGATCAGGAGAAGTACATCCCCGAAGAGATCTTTCTTCATTTCTCT 560
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    363 GGCCTTGGAGCTCCTCGTCGATCAGGAGAAGTACATCCCCGAAGAGATCTTTCTTCATTTCTCT 422

Qy    561 GGGCCATGATGAGGAGAGTCATCATCAGGGAGGAGGGCCTCAGGAGGATCTCAGGACCCCTGCTACAGTC 620
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    423 GGGCCATGATGAGGAGAGTCATCATCAGGGAGGAGGGCCTCAGGAGGATCTCAGGACCCCTGCTACAGTC 482

Qy    621 AAGGGGCGGCGTCCAGCTAGC 638
          |||||||||||||||||||||
Db    483 AAGGGGCGGCGTCCAGCTAGC 500
```

```
RESULT 3
AAX40258
ID   AAX40258 standard; cDNA; 319 BP.
XX
AC   AAX40258;
XX
DT   16-JUN-1999 (first entry)
XX
DE   Human secreted protein 5' EST SEQ ID NO:45.
XX
KW   Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW   forensic; gene therapy; chromosome mapping; signal peptide;
KW   upstream regulatory sequence; cytokine activity; cell proliferation;
KW   differentiation; haematopoiesis regulation; tissue growth regulation;
KW   reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW   thrombolytic; anti-inflammatory; tumour inhibition; ds.
XX
OS   Homo sapiens.
XX
PN   WO9906439-A2.
XX
PD   11-FEB-1999.
XX
PF   31-JUL-1998; 98WO-IB01233.
XX
PR   01-AUG-1997; 97US-0904468.
XX
PA   (GEST ) GENSET.
XX
PI   Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR   WPI; 1999-153700/13.
DR   P-PSDB; AAY11540.
XX
PT   New nucleic acids encoding human secreted proteins - obtained from
```

PT cDNA libraries derived from liver, lung, large intestine, colon,
PT   thyroid and pancreas tissue
XX
PS Claim 1; Page 175-176; 398pp; English.
XX
CC AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY11533 to
CC AAY11679, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 319 BP; 75 A; 86 C; 90 G; 68 T; 0 other;

Query Match   17.7%; Score 267; DB 20; Length 319;
Best Local Similarity   99.6%; Pred. No. 1.2e-74;
Matches 278; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy   1 ATGGCTCAGCGGGGTGCGTTGCGTGTCGGCCCTGGGCTATGCTGCTCCTCAGTTTTCCCT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 ATGGCTCAGCGCGGTGCGTTTGCGTGTCGGCCTATGCTGCTCCTCAGTTTTCCCT 100

Qy   61 ACCGTCTCCAGATCGATCGGGCCCGAGGAGCGGGGAGCATCAAAGGGCGTCGCCGAATCCCTT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ACCGTCTCCAGATCGATCGGGGCCCGAGGAGCGGGGAGCATCAAAGGGCGTCGCGAAATCCCT 160

Qy   121 TCTCAGTTCAGCAAAAGAGGAACGCGTCGCGAGTGAAAAGGCGTGCCATCCAG 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 TCTCAGTTCAGCAAAAGAGGAACGCGTCGCGATGAAAAGGAGCGTGCCATCCAG 220

Qy   181 ATTCCACACAGTGACTTTTAGCTCTGAGAAGTTCCAATACTACAGCCCTGGCTGAGT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGG- 279

Qy   241 AAATACACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACC 279
          ||||||||||||||||||||||||||||||||||||||||||
Db   280 AAATACACATTCATAAAGTCTTTCCTACACAGTGGTCAGCACC 318

PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US10836.
XX
PR 30-MAY-2000; 2000US-206132P.
PR 29-AUG-2000; 2000US-228716P.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shimkets RA, Leach MD;
XX
PS WPI; 2002-1063308/14.
DR P-PSDB; ABP09238.
XX
XX
CC Novel human polypeptides and polynucleotides useful for diagnosing,
CC preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders
XX
XX Disclosure; SEQ ID 18457; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
CC disorder in humans, and in the manufacture of a medicament for treating a
CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC sequences can be used for diagnosis or gene therapy. ORFX sequences can be used in the
CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC osteoarthritis, neurodegenerative disorders, disorders related to organ
CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC storage disease, various immune deficiencies and disorders, infectious
CC diseases, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 348 BP; 77 A; 102 C; 89 G; 79 T; 1 other;

Query Match   16.6%; Score 251; DB 24; Length 348;
Best Local Similarity   82.7%; Pred. No. 1.7e-69;
Matches 287; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Qy   176 TCCAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGT 235
          ||| |||||||||||||||| |||||||||||||||||||||||||||||||||| ||||
Db   2 TCCGGATTCCCACAGTGTCTTTCAGCCACGAGAATCCAACACCACAGCCCTTGCTGAGT 61

Qy   236 TCGGAAAATACAITCATAAAGTCTTTCCTACACAGTGGTCAGCACCAGCTTTATCCAGCATG 295
          | ||| ||||||| ||||| |||| |||||||||||||||||||||||||||||| |||
Db   62 TTGGAGAATATATCCGAAAAGCCTTTCCTACAGTGTTCCACAGCAACCTTATCCAGCACAG 121

Qy   296 AAGTCGTGGAAGAGTATAGCCACCCTGTTCACTATCCANAGGCTGGACCCCAGCGTTGCAGC 355
          |||||||| | ||||||||||||||||||| ||| ||||||||||||||||||||||| ||
Db   122 AAGTCGTGGGGAAGATATAGCCACCCTGTTCACCGTCCGGAGTCGGACACCCAGCGTTGCAGC 181

Qy   356 CCTACCTGCTGATGGCCTCCACTTTGATCGTGGTCCCCTGAAGGNAAGGTGGGAGGGTGC 415
          ||||||||||||||||||||||| ||||||||||| ||||||||||||||||||||||||
Db   182 CCTACATGCTGATGGCTCCACTTCGACGTGGTTCCTCGCCTCTGAAGAAGGATAGGGAGGTGC 241

Qy   416 CCCCATTCTCTCTGGTTGGGCGTGCATGGCCTGGCGTCATCTATGGTTCGGCGGCCACACTGGACGACA 475
          ||||||||||||||||||||||||||||||||||||||||||||| ||||| ||||||||| ||
Db   242 CCCCGTTCTTCAGCGCCTGGAGCAAAAATGGCTCATTCCATGGCCGCGGGGGTGCCGCGGGGACAAA 301

Qy   476 AGAACTCTGTGATGGCAITACTGCAGGCCTTGGGAGCTCCTGCTGATC 522
          ||||||||||||||||||||||||||||||||||||||||||||||||
          (additional sequence continues)

ID ABN24990 standard; cDNA; 348 BP.
XX
AC ABN24990;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX polynucleotide sequence SEQ ID NO:18457.
XX

RESULT 4
ABN24990
ID ABN24990 standard; cDNA; 348 BP.
XX
AC ABN24990;
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis; gene; ss.
XX
OS Homo sapiens.
XX
PN WO20019523-A2.
XX

RESULT 5

```
Db   302 AAAACTCTGTGATGGCAGTCCTGCAGGCTTTGGAGCTCCTGTTGATC 348
         |||||||||||||| |||||||||| ||||||||||||||||| |||||
Qy   302 AAAACTCTGTGATGGCAGTCCTGCAGGCTTTGGAGCTCCTGTTGATC 348
```

AAI84929
ID   AAI84929 standard; cDNA; 362 BP.
XX
AC   AAI84929;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polynucleotide SEQ ID NO 4989.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation; ss.
XX
OS   Homo sapiens.
XX
PN   WO200164835-A2.
XX
PD   07-SEP-2001.
XX
PR   26-FEB-2001; 2001WO-US04927.
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   P-PSDB; AAO04998.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 1; SEQ ID NO 4989; 1399pp + Sequence listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 362 BP; 72 A; 101 C; 109 G; 79 T; 1 other;

Query Match   16.2%; Score 245; DB 22; Length 362;
Best Local Similarity   79.7%; Pred. No. 1.4e-67;
Matches 342; Conservative   0; Mismatches   0; Indels 87; Gaps   1;
```

```
Qy     1 ATGGCTCAGCGGTGCGTCGGTTGGCGTCGCCCTCGGTGGCTATGCTGCTGCTCTAGTTTTTCCCT  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    20 ATGGCTCAGCGGTGCGTTGCGTTTGCGTGCTGCCCCTCGGTGGCTATGGCTGCTCCTAGTTTTTCCCT  79

Qy    61 ACCGTCTCCAGATCGATGGGGCCCGAGAGACGGGGGAGCATCAAAGGCGGTTCGGCGAAATCCCT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    80 ACCGTCTCCAGATCGATGGGGCCCGAGAGACGGGGGAGCATCAAAGGCGGTTCGGCGAAATCCCT 139

Qy   121 TCTCAGTTCAGCAAAGAGGAACGCGTCCGATGAAAAGAGGAGGCGCTGAAAAGTGCCATCCAG 180
```

RESULT 6

```
Db   140 TCTCAGTTCAGCAAAGAGGAACGCGTCCGATGAAAAGAGGAGGCGCTG------------- 184
         |||||||||||||||||||||||||||||||||||||||||||||||
Qy   181 ATTCCAAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTAGTCCGA ------- 184
```

AAI73675
ID   AAI73675 standard; DNA; 51 BP.
XX
AC   AAI73675;
XX
DT   09-NOV-2001 (first entry)
XX
DE   Human silent SNP containing nucleic acid SEQ:616.
XX
KW   Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW   protein therapy; vaccine; probe; diagnostic assay; detection;
KW   quantitation; restorative therapy; polymorphic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200140521-A2.
XX
PD   07-JUN-2001.
XX
PR   30-NOV-2000; 2000WO-US32758.
PR   30-NOV-1999; 99US-0168138.
PR   29-NOV-2000; 2000US-0726173.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2001-356160/37.
XX
PT   Polymorphic nucleic acid sequences, useful in genetic testing and
PT   therapy -
XX
PS   Claim 1; Page 243; 2653pp; English.
XX
CC   AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC   sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC   AAM53114 to AAM53329 represent peptides related to human polymorphic
CC   polynucleotide sequences. The sequences can be used in gene and protein
CC   therapy, and in vaccine production. (I) and the polypeptides encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression of polymorphic polypeptides.
CC   For example, (I) may be used to treat disorders by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patients own production of polypeptide. Additionally, (I) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids
```

```
Db   185 AAATACATTCATAAAGTCTTTTCCTACAGTGGTCACCACCAGCTTTATCCAGCATGAAGTC 232
Db   185 -------- AAAGTCTTTTCCTACAGTGGTCACCACCAGCTTTATCCAGCATGAAGTC 232
Qy   241 AAATACATTCATAAAGTCTTTTCCTACAGTGGTCACCACCAGCTTTATCCAGCATGAAGTC 300

Db   233 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCTCGGACCCCAGCTTGCAGCCCTAC 292
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGCTCGGACCCCAGCTTGCAGCCCTAC 360

Db   293 CTGCTGATGGCTCACTTTGATGTGGTTGCCGTTTCGTGCGCCCA 352
         |||||||||||||||||||||||||||||||||||||||||||
Qy   361 CTGCTGATGGCTCACTTTGATGTGGTTGCCGTGGGAGGTGCCCCA 420

Db   353 TTCTCTCTGGG 361
         |||||||||||
Qy   421 TTCTCTCTGG 429
```

CC   In samples, and therefore which patients may be in need of restorative
CC   therapy. The polypeptides encoded by (I) may be used as antigens in the
CC   production of antibodies specific for polymorphic polypeptides. The
CC   antibodies may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of polymorphic polypeptides in samples.
XX
SQ   Sequence 51 BP; 12 A; 12 C; 18 G; 9 T; 0 other;

Qy 432 GGAGCGTGATGGCGTCATCTATGGTCGGGCCACACTGGACGACAAGAACTC 482
        |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 GGAGCCTGATGGCGTCATCTATGGTCGGGCCACACTGGACGACAAGAACTC 51

RESULT 7
AAQ33105
ID   AAQ33105 standard; cDNA to mRNA; 1221 BP.
XX
AC   AAQ33105;
XX
DT   06-MAY-1993 (first entry)
XX
DE   Pig coding sequence for aminoacylase I.
XX
KW   AmA; production; yield; recombinant; ss.
XX
OS   Sus scrofa.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1221
FT   /*tag= a
XX
PN   JP04330279-A.
XX
PD   18-NOV-1992.
XX
PF   22-MAR-1991; 91JP-0081136.
XX
PR   22-MAR-1991; 91JP-0081136.
XX
PA   (TAKI) TAKARA SHUZO CO LTD.
XX
DR   WPI; 1993-003494/01.
DR   P-PSDB; AAR30458.
XX
PT   Polypeptide with amino:acylase (I) activity - has specific base
PT   sequence and is used for preparing probe, primer and antibody
XX
PS   Claim 2; Page 9; 14pp; Japanese.
XX
CC   Pig aminoacylase I cDNA is cloned from pig kidney poly (A) mRNA
CC   giving clones lambda pKamA -1 to -10. The longest insert fragment in
CC   lambda pKamA-10 (1.4kb) was isolated and sequenced. The sequence
CC   may be used to transform Saccharomyces cerevisiae to produce
CC   aminoacylase I recombinantly.
CC   See also AAQ33106.
XX
SQ   Sequence 1221 BP; 241 A; 344 C; 381 G; 255 T; 0 other;

Query Match   3.4%; Score 50.8; DB 14; Length 1221;
Best Local Similarity 51.9%; Pred. No. 4e-05;
Matches 140; Conservative   0; Mismatches 127; Indels   3; Gaps   1;

Qy 318 CCTGTGTCACTATXCCAAGGCTCGGACCCCTAGCTTGCAGCCCTACCTGATGGGCTCACTT 377
        | ||| ||| | || ||||| | ||||| ||||||||||| || || ||||| ||
Db 183 CGTGCGTGACCTGGCCCGGCGCACCAAACCCACCACTCTTCCTCCCATCTTCTTGCCACAC 242

Qy 378 TGATGTGGTGCTGCCCGTGCCCCCTGAAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTT---GGA 434
        |||||||||||| ||| ||| ||| |||||
Db   ...

Db 243 AGATGTGGTGCTGTCTTCTTCAAGGAGACAATTGGAGTCATTGACCCCTTGAGGGCTTCAAGGA 302
Qy 435 GCCTGAGTGGCGTCATCTATGGGTCGGGCGGGCCACACTGGACAGCACAAGAACTCTCGATGGCATT 494
Db 303 TGCAGAGTAGGCTACACTCTATTGCCAGGGCGGCCCCAGCACATGAAGAGTGCGTGCAGCATCCAGTA 362
Qy 495 ACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGGTACATCCCCGGAAAGATCTTTCTTCAT 554
Db 363 CCTGGAGGCGTGAGGAGGGTTGAGGGCCACCAATTTCCCAGAGAAGCCATCCACAT 422
Qy 555 TTCTCTGGGCCATGAGTGAGGAGGTCATCAGG 584
Db 423 GACCTTTGTTGCCAGGATGAGGAGGTTGGGAGG 452

RESULT 8
AAI73672
ID   AAI73672 standard; DNA; 51 BP.
XX
AC   AAI73672;
XX
DT   09-NOV-2001 (first entry)
XX
DE   Human silent SNP containing nucleic acid SEQ:613.
XX
KW   Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW   protein therapy; vaccine; probe; diagnostic assay; detection;
KW   quantitation; restorative therapy; polymorphic; ds.
XX
OS   Homo sapiens.
XX
PN   WO200140521-A2.
XX
PD   07-JUN-2001.
XX
PR   30-NOV-2000; 2000WO-US32758.
PR   30-NOV-1999; 99US-0168138.
PR   29-NOV-2000; 2000US-0726173.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2001-356160/37.
XX
PT   Polymorphic nucleic acid sequences, useful in genetic testing and
PT   therapy -
XX
PS   Claim 1; Page 242; 2653pp; English.
XX
CC   AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC   sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC   AAM53114 to AAM53329 represent peptides related to human polymorphic
CC   polynucleotide sequences. The sequences can be used in gene and protein
CC   therapy, and in vaccine production. (I) and the polypeptides encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression of polymorphic polypeptides.
CC   For example, (I) may be used to treat disorders by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patients own production of polypeptide. Additionally, (I) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids
CC   in samples, and therefore which patients may be in need of restorative
CC   therapy. The polypeptides encoded by (I) may be used as antigens in the
CC   production of antibodies specific for polymorphic polypeptides. The
CC   antibodies may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of polymorphic polypeptides in samples.
XX
SQ   Sequence 51 BP; 8 A; 9 C; 19 G; 15 T; 0 other;

Query Match   3.3%;   Score 49.4; DB 22; Length 51;
Best Local Similarity   98.0%;   Pred. No. 1.5e-05;
Matches 50; Conservative   0; Mismatches 1; Indels   0; Gaps   0;

Qy   420 ATTCTCTGGGTTGGAGCGTGATGGGCGTCATCTATGGTTGGGGCACACTGGA 470
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 ATTCTCTGGGTTGGAGCGTGATGGGCGTCATCTATGGTTGGGGCACACTGGA 51

RESULT 9
AAI73674
ID   AAI73674 standard; DNA; 51 BP.
XX
AC   AAI73674;
XX
DT   09-NOV-2001 (first entry)
XX
DE   Human silent SNP containing nucleic acid SEQ:615.
XX
KW   Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW   protein therapy; vaccine; probe; diagnostic assay; detection;
KW   quantitation; restorative therapy; polymorphic; ds.
XX
OS   Homo sapiens.
XX
XX
PN   WO200140521-A2.
XX
PD   07-JUN-2001.
XX
PF   30-NOV-2000; 2000WO-US32758.
XX
PR   30-NOV-1999;   99US-0168138.
PR   29-NOV-2000; 2000US-0726173.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach M;
XX
DR   WPI; 2001-356160/37.
XX
PT   Polymorphic nucleic acid sequences, useful in genetic testing and
PT   therapy -
XX
PS   Claim 1; Page 242; 2653pp; English.
XX
CC   AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC   sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC   AAM53114 to AAM53329 represent peptides related to human polymorphic
CC   polynucleotide sequences. The sequences can be used in gene and protein
CC   therapy, and in vaccine production. (I) and the polypeptides encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases
CC   associated with inappropriate expression of polymorphic polypeptides.
CC   For example, (I) may be used to treat disorders by rectifying mutations
CC   or deletions in a patient's genome that affect the activity of
CC   polypeptides by expressing inactive proteins or to supplement the
CC   patients own production of polypeptide. Additionally, (I) and its
CC   complementary sequences may also be used as DNA probes in diagnostic
CC   assays to detect and quantitate the presence of similar nucleic acids
CC   in samples, and therefore which patients may be in need of restorative
CC   therapy. The polypeptides encoded by (I) may be used as antigens in the
CC   production of antibodies specific for polymorphic polypeptides. The
CC   antibodies may also be used to down regulate expression and activity.
CC   The antibodies may also be used as diagnostic agents for detecting the
CC   presence of polymorphic polypeptides in samples.
XX
SQ   Sequence 51 BP; 12 A; 11 C; 18 G; 10 T; 0 other;

Query Match   3.3%;   Score 49.4; DB 22; Length 51;
Best Local Similarity   98.0%;   Pred. No. 1.5e-05;
Matches 50; Conservative   0; Mismatches 1; Indels   0; Gaps   0;

Qy   432 GGAGCGTGATGGGCGTCATCTATGGTTCGGGCGCACACTGGACGACGAAGAACTC 482
              ||||||||||||||||||||||||||||| ||||||||||||||||||||||||||

Db   1 GGAGCGTGATGGGCGTCATCTATGGTTGGGGCACACTGGACGACGAAGAACTC 51

RESULT 10
AAQ33106
ID   AAQ33106 standard; cDNA to mRNA; 1224 BP.
XX
AC   AAQ33106;
XX
DT   06-MAY-1993 (first entry)
XX
DE   Human coding sequence for aminoacylase I.
XX
KW   AmA; production; yield; recombinant; ss.
XX
OS   Homo sapiens.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..1224
FT   /*tag= a
XX
PN   JP04330279-A.
XX
PD   18-NOV-1992.
XX
PF   22-MAR-1991; 91JP-0081136.
PR   22-MAR-1991; 91JP-0081136.
XX
PA   (TAKI) TAKARA SHUZO CO LTD.
XX
DR   WPI; 1993-003494/01.
DR   P-PSDB; AAR30459.
XX
PT   Polypeptide with amino:acylase (I) activity - has specific base
PT   sequence and is used for preparing probe, primer and antibody
XX
PS   Claim 2; Page 11; 14pp; Japanese.
XX
CC   Human aminoacylase I cDNA was determined. The sequence
CC   may be used to transform Saccharomyces cerevisiae to produce
CC   aminoacylase I recombinantly.
CC   See also AAQ33105.
XX
SQ   Sequence 1224 BP; 250 A; 353 C; 368 G; 253 T; 0 other;

Query Match   3.3%;   Score 49.2; DB 14; Length 1224;
Best Local Similarity   51.9%;   Pred. No. 0.00013;
Matches 136; Conservative   0; Mismatches 123; Indels   3; Gaps   1;

Qy   318 CCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTT 377
              | ||||| | |||||||| | |||| |||||| || |||| || ||| |||| |||||
Db   183 CGTGTTGACCTGGCCAGGCACCAAACCCTACTCTCCTCCCATCTTCTGCTGGGTT---GGA 242

Qy   378 TGAWGTGTGGTGCCTGCCCCTGAAGAAGCTGGGAGGTGCCCCTCCCATTCTGGGTT---GGA 434
              ||||||| ||||| | ||||| | ||| ||| ||| ||| ||| || ||| | ||| |||
Db   243 GGAWGTGGTGCCCTGTCTTCAAGGAACATTGGAGTCACGACCCCTTGAGGCCTTCAAGGA 302

Qy   435 GCGTGAWGGCGTCATCTRATGGTTCGGGCGCACACTGGACGACGAAGAACTCTGTGWGATGGCATT 494
              || | || || ||| | ||| || ||| || | ||| ||||| ||||| |||| | ||||
Db   303 TTCTGAGGGCTACATCTATGCCAGGGGTGCCCAGGACATGAAGTGCCGTCAGCATCAGTA 362

Qy   495 ACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGRAAGAGTCTTTTCTTCAT 554
              ||| ||| ||||| ||| | ||| |||| || || |||| | ||| ||| || ||| |||
Db   363 CCTGGAAGCTGTGAGGAGGCTGAAGGGTGGAGGGCCACCGGGTTCCCAGAAAACCATCCACAT 422

Qy   555 TTCTCTGGGCCCATGATGAGGAG 576
              || | || |||||| || |||||
Db   423 GACCTTTGTGCCCTGATGAGGAG 444

RESULT 11
ABK83807

ID   ABK83807 standard; cDNA; 1415 BP.
XX
AC   ABK83807;
XX
DT   14-AUG-2002 (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #378.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
OS   Homo sapiens.
XX
PN   WO20022B999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US30821.
PR   03-OCT-2000; 2000US-237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression
PT   of genes associated with granulocyte activation, which serves as
PT   diagnostic markers that is useful for monitoring disease states and
PT   drug toxicity -
XX
PS   Claim 1; SEQ ID NO. 378; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (GS) identified by
CC   DNA chip analysis as given in the specification, and comparing
CC   the expression level to an expression level in an unactivated
CC   GC, where differential expression of Gs is indicative of GCA.
CC   Also included are modulating (M2) GA by contacting GC with an agent
CC   that alters the expression of at least one gene in Gs; (2) screening (M3)
CC   for an agent capable of modulating GCA or an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease using the
CC   gene expression profile; (3) detecting (M4) an inflammation (especially
CC   chronic) in a tissue, an allergic response in a subject, exposure of a
CC   subject to a pathogen or sterile inflammatory disease, by detecting the
CC   level of expression in a sample of the tissue of gene(s) from Gs, where
CC   the level of expression of the gene is indicative of inflammation;
CC   (4) treating (M5) an inflammation (especially chronic) or in a tissue,
CC   an allergic response in a subject, exposure of a subject to a pathogen
CC   or sterile inflammatory disease, by contacting a tissue having
CC   inflammation with an agent that modulates the expression of gene(s)
CC   from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for
CC   modulating GA; M3 is useful for screening an agent capable of modulating
CC   GCA preferably in an inflammation in a tissue; M4 is useful for
CC   detecting an inflammation (especially chronic) in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease (e.g. psoriasis, rheumatoid arthritis,
CC   glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal
CC   reperfusion injury, ARDS, adult respiratory distress syndrome,
CC   inflammatory bowel disease, Crohn's disease, ulcerative colitis,
CC   periodontal disease; also bacterial infection, viral infection,
CC   parasitic infection, protozoal infection, fungal infection and M5 is
CC   useful for treating one of the above conditions. The present
CC   sequence represents a gene differentially expressed in granulocytes.
CC   Note: The sequence data for this patent did not form part

CC   of the printed specification, but was obtained in electronic
CC   format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 1415 BP; 297 A; 416 C; 416 G; 286 T; 0 other;

Query Match   3.3%; Score 49.2; DB 24; Length 1415;
Best Local Similarity   51.9%; Pred. No. 0.00014;
Matches 136; Conservative   0; Mismatches 123; Indels   3; Gaps 1;

Qy   318 CCTGTTCACTATCCCAAGGCGTCGGACCCCCAGCCTACCTGCTGCAGCCTCAGCTTGATGGCTGACTT 377
          || |||| || | |||||| || |||| ||||| | || ||||||| ||| ||| ||||
Db   244 CGTGTTGACCTGGCCCAGGCCACCACCACCCTACACTCTCCATCTTGCTCAAACTCCCACAC 303

Qy   378 TGATGTCGTGCCTGCCCCCTGAAGAAGGACTGGGAGGTGCCCCATTCTTCTGGGGTT---GGA 434
          ||||||||||| |||| | ||||||||||||||| || ||||||||||||| ||| |||
Db   304 GGATGTGGTGCCTGCCTGCTGTTCAAGGAACAATTGGAGTCACGACAAGAACTGTGTGATGGCATT 363

Qy   435 GCGTGATGGCGCCTTCATCTTATGTGCGGGCGCACACTGGACGACAAGAAACTCTGTGTGATGGCATT 494
          ||| |||| ||| ||||||||| | ||| || |||| ||| || | ||| || |||| |||| |||
Db   364 TTCTGAGGGCTACACATCTTATGCCAGGGGGTGCCCAGGACATGAAGTGCGTGCAGCATCCAGTA 423

Qy   495 ACTGCAGGCCCTTGGAGCTCCTCGCCTGATCAGGAAGTACATCCCCGAAGATCTTTCTTCAT 554
          ||||| ||| || |||| | | |||||||| || | ||| |||||||| |||||| |||
Db   424 CCTGGAAGCTGTGGTGGAGGGCTGAAGGTGGAGGCCCACCGGTTCCCCAGAACCATCCACAT 483

Qy   555 TTCTCTCGGGCCCATGATTGAGGAG 576
          | || |||||||||||||||||||
Db   484 GACCTTTGGTGCCTGATGAGGAGGAG 505

RESULT 12
AAI73673
ID   AAI73673 standard; DNA; 51 BP.
XX
AC   AAI73673;
XX
DT   09-NOV-2001 (first entry)
XX
DE   Human silent SNP containing nucleic acid SEQ:614.
XX
KW   Human; single nucleotide polymorphism; SNP; genome; gene therapy;
KW   protein therapy; vaccine; probe; diagnostic assay; detection;
KW   quantitation; restorative therapy; polymorphic; ds.
OS   Homo sapiens.
XX
PN   WO20014D521-A2.
XX
PD   07-JUN-2001.
XX
PF   30-NOV-2000; 2000WO-US32758.
PR   30-NOV-1999; 99US-0168138.
PR   29-NOV-2000; 2000US-0726173.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2001-356160/37.
XX
PT   Polymorphic nucleic acid sequences, useful in genetic testing and
PT   therapy -
XX
PS   Claim 1; Page 242; 2653pp; English.
XX
CC   AAI73060 to AAI79867 represent isolated human polymorphic polynucleotide
CC   sequences (I), which contain single nucleotide polymorphisms (SNPs).
CC   AAM53114 to AAM53329 represent peptides related to human polymorphic
CC   polynucleotide sequences. The sequences can be used in gene and protein
CC   therapy, and in vaccine production. (I) and the polypeptides encoded by
CC   them may be used in the prevention, diagnosis and treatment of diseases

Right block (RESULT 14)

CC Sequences ABP31028-ABP35561 represent 4534 novel human proteins
CC designated ORF (open reading frame) 1-4534, and sequences ABN75054-
CC ABN79587 represent cDNAs encoding them. The invention also encompasses
CC polypeptides at least 80% identical to the ORF1-ORF4534 (collectively
CC referred to as ORFX) proteins, polynucleotides at least 85% identical to
CC the ORFX nucleic acid sequences, vectors and host cells comprising ORFX
CC polynucleotides, the recombinant production of ORFX proteins, antibodies
CC specific for ORFX proteins, methods of detecting ORFX polynucleotides and
CC polypeptides, methods of screening for modulators of ORFX expression or
CC activity, and methods of screening individuals for a predisposition to an
CC ORFX-associated disorder. The ORFX proteins of the invention have a wide
CC range of biological activities, such as cytokine, cell proliferation,
CC cell differentiation, immune modulation, haematopoiesis regulation,
CC tissue growth, angiogenesis, activin or inhibin activity, chemotactic/
CC chemokinetic activity, haemostatic activity, tumour inhibition activity,
CC receptor/ligand, antiinflammatory activity,
CC and antiinfective activity, and may also be involved in the determination
CC of bodily characteristics, fertility and behaviour. ORFX proteins,
CC nucleic acids and antibodies may be used in the treatment of cancers,
CC other proliferative disorders such as psoriasis and benign tumours,
CC neurological disorders such as epilepsy and Alzheimer's disease,
CC cardiovascular diseases, immune system disorders, disorders related to
CC organ transplantation, disorders of tissue growth and regeneration,
CC diseases such as diabetes mellitus, hypothyroidism, and cholesterol ester
CC storage disease, and infectious diseases caused by viral, bacterial,
CC fungal and other pathogens. ORFX nucleic acids may also be used as a
CC source of primers and probes, in the detection of ORFX genomic sequences
CC or transcripts, in the identification and cloning of homologous
CC sequences, in genetic diagnosis, and in forensic biology. The ORFX
CC nucleic acids may additionally be used to produce transgenic animals
CC which may be useful for studying the function and/or activity of ORFX
CC protein, and in drug screening. The ORFX proteins may also be used as
CC immunogens to generate specific antibodies, which are useful in the
CC diagnosis, treatment and monitoring of ORFX-associated diseases.
XX
SQ Sequence 381 BP; 111 A; 93 C; 92 G; 82 T; 3 other;

Qy   Query Match 3.1%; Score 46.2; DB 24; Length 381;
       Best Local Similarity 94.1%; Pred. No. 0.0057; Matches 48; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1040 TCAAGTTCAAATGTCATCCCCCCAGTGGCCCAGGCCACAGTCAACTTCCGGA 1090
         || |||||||||||||||| |||||||||||||||||||||||||||| |||
Db   331 TCTAGTTCAAATGTCATCCACCCAGTGGCCCAGGCCACAGTCAACTTCGGGA 381

RESULT 14
ABL60776
ID ABL60776 standard; cDNA; 1445 BP.
XX
AC ABL60776;
XX
DT 10-SEP-2002 (first entry)
XX
DE cDNA encoding an enzyme similar to human aminoacylase-1 (ACY-1).
XX
KW Aminoacylase-1; ACY-1; metalloprotein; cytosolic enzyme; human; gene;
KW cytostatic; therapeutic; cancer therapy; ss.
XX
OS Homo sapiens.
XX

Key	Location/Qualifiers
FH	
FT 5'UTR	1..116
FT	/tag= a
FT CDS	117..1238
FT	/tag= b
FT 3'UTR	1239..1445
FT	/tag= c

XX
PN US6387661-B1.
XX
PD 14-MAY-2002.
XX

Left block (RESULT 13)

CC associated with inappropriate expression of polymorphic polypeptides.
CC For example, (I) may be used to treat disorders by rectifying mutations
CC or deletions in a patient's genome that affect the activity of
CC polypeptides by expressing inactive proteins or to supplement the
CC polypeptides own production of polypeptide. Additionally, (II) and its
CC complementary sequences may also be used as DNA probes in diagnostic
CC assays to detect and quantitate the presence of similar nucleic acids
CC in samples, and therefore which patients may be in need of restorative
CC therapy. The polypeptides encoded by (I) may be used as antigens in the
CC production of antibodies specific for polymorphic polypeptides. The
CC antibodies may also be used to down regulate expression and activity.
CC The antibodies may also be used as diagnostic agents for detecting the
CC presence of polymorphic polypeptides in samples.
XX
SQ Sequence 51 BP; 9 A; 9 C; 18 G; 15 T; 0 other;

Qy   Query Match 3.2%; Score 47.8; DB 22; Length 51;
       Best Local Similarity 96.1%; Pred. No. 5e-05; Matches 49; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   420 ATTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTGTCGCGGGGGCCACACTGGA 470
         |||||||||||||||||||| |||| | |||||||||||| || ||||||||||||
Db   1 ATTCTCTGGGTTGGAGCGTGANGGCANCATCTATGGTGTCGGGGCACACTGGA 51

RESULT 13
ABN75263
ID ABN75263 standard; cDNA; 381 BP.
XX
AC ABN75263;
XX
DT 08-JUL-2002 (first entry)
XX
DE Human ORF210 cDNA, SEQ ID NO:419.
XX
DE Human; ORF; open reading frame; ORFX; drug screening; diagnosis;
KW disease monitoring; cytokine; cell proliferation; cell differentiation;
KW immune modulation; haematopoiesis regulation; tissue growth;
KW angiogenesis; activin; inhibin; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; tumour inhibition; bodily characteristic; fertility;
KW behaviour; cancer; proliferative disorder; neurological disorder;
KW cardiovascular disease; immune system disorder; organ transplantation;
KW tissue growth disorder; tissue regeneration disorder; diabetes mellitus;
KW hypothyroidism; cholesterol ester storage disease; infection; vulnerary;
KW vasotropic; antipsoriatic; antidiabetic; cytostatic; nootropic;
KW cardiant; hypotensive; antithyroid; antiinflammatory; immunomodulator;
KW neuroprotective; antiatherosclerotic; anticoagulant; thrombolytic;
KW dermatological; analgesic; virucide; antibacterial; fungicide; gene; ss.
OS Homo sapiens.
XX
PN WO200190366-A2.
XX
PD 29-NOV-2001.
XX
PA (CURA-) CURAGEN CORP.
PI Leach MD, Shimkets RA;
XX
DR WPI: 2002-106200/14.
DR P-PSDB; ABP31237.
XX
PF 24-MAY-2001; 2001WO-US17076.
PR 24-MAY-2000; 2000US-206690P.
XX
OS Homo sapiens.
XX
CC Novel human polypeptides and polynucleotides useful for diagnosing,
CC preventing and treating cardiovascular disease, neurodegenerative,
CC hyperproliferative disorders and disorders related to organ
CC transplantation
XX
PS Claim 1; Page 373; 2508pp; English.
XX

RESULT 15

ID ABK30982 standard; cDNA; 663 BP.
XX
AC ABK30982;
XX
DT 23-APR-2002 (first entry)
XX
DE Plant dwarfing/stunting related cDNA seq ID 396.
XX
KW plant; ss; dwarfism; stunting; EST; expressed sequence tag;
KW transgenic plant; plant metabolism.
XX
OS Planta.
XX
PN WO200208410-A2.

PF 23-MAR-2001; 2001US-0814951.
XX
PR 23-MAR-2001; 2001US-0814951.
XX
PA (PEKE) PE CORP NY.
XX
PI Shao W, Yan C, Di Francesco V, Beasley EM;
XX
DR WPI; 2002-478443/51.
DR P-PSDB; AB808102.
XX
PS Claim 5; Fig 1A; 43pp; English.
XX
PT Isolated nucleic acid molecules encoding enzymes similar to human
PT aminoacylase-1, useful as a drug target and diagnostic marker for
PT cancers e.g. T cell leukemias and ovary, brain or lung cancers -

CC The invention relates to an isolated nucleic acid molecule encoding
CC enzymes similar to human aminoacylase-1 (ACY-1) (EC 3.5.1.14) (a
CC metalloprotein cytosolic enzyme). The ACY-1 similar polynucleotide and
CC encoded peptide sequences can be used as models for the development of
CC human therapeutic targets, aid in the identification of therapeutic
CC proteins, and serve as targets for the development of human therapeutic
CC agents that modulate enzyme activity in cells and tissues that express
CC the enzyme. ACY-1 has been found to be expressed in humans in the
CC placenta, T cells from T cell leukemia, ovary, brain, lung and leukocyte,
CC and therefore may be a drug target for cancer therapy and act as a
CC diagnostic marker for these cancers. The present sequence represents a
CC cDNA encoding an enzyme similar to human aminoacylase-1 (ACY-1).
XX
SQ Sequence 1445 BP; 375 A; 410 C; 391 G; 269 T; 0 other;

Query Match   3.0%; Score 46; DB 24; Length 1445;
Best Local Similarity 51.6%; Pred. No. 0.0015;
Matches 131; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

```
Qy   326 CTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCGTCATGGCTCACTTTGATGTGG 385
         ||   ||| |  ||||  ||| | |||||| ||| ||| |||||||||||||||
Db   202 CTGACTAPGGCACCAACCCTACACTCTCCCATCTTGCTCAACTCCCACACGGAPGTGG 261

Qy   386 TGCCTGCCCCTGAAGAAGGCTGGGGAGGTGCCCCCAATTCTCTGGGTT---GGAGCGTGATG 442
         ||   |  |  |||  ||||   ||  | |||  |     | |  ||||   ||  |||||
Db   262 TGCCTGTCTTCAAGGAACATTGGAGTCACGAGCCCTTTGAGGCCTTCAAGGATTCTGAGG 321

Qy   443 GCGGTCATCTATGGTCGGGGGCACACTGGACGACAAGAAACTCGTGTGATGGGCAATTACTGCAGG 502
         ||  | |||  | | ||||| |  |||| ||| || || |||||||||||||
Db   322 GCTACAACTATCCAGGGGGTCCCCAGGACATGAAGTGCGTCAGCATCCAGTACCTGGAAG 381

Qy   503 CCTTGGAGCTCCTGCTCGATCAGGAAGTACAATCCCCGGAAGATCTTTTCTTGCATTTCTCTCGG 562
         | ||| ||  |  |  | |  | |  || |  |||  ||||| ||  || ||  | |||
Db   382 CTGTGAGGAGGCTGAAGGTGGAGGGCCACCCGGTTCCCAGAGAACCATCCACATGACCTTTG 441

Qy   563 GCCATGATGAGGAG 576
         | |||||||||||
Db   442 TGCCTGATGAGGAG 455
```

XX
PD 31-JAN-2002.
XX
PF 20-JUL-2001; 2001WO-US23120.
XX
PR 20-JUL-2000; 2000US-2198009P.
PR 20-JUL-2000; 2000US-2198010P.
XX
PA (DOWC) DOW CHEM CO.
PA (REDD/) REDDY S A.
PA (LARR/) LARRINUA M I.
PA (RUEG/) RUEGGER M.
PA (WEGL/) WEGLARZ T.
PA (BLAK/) BLAKESLEE B.
PA (ORIE/) ORIEDO V B J.
PA (SAVI/) SAVICKAS J P.
PA (MCCR/) MCCRERY A D.
PA (MILL/) MILLER A B.
PA (GACH/) GACHOTTE D.
PA (GROS/) GROSLEY R.
PA (PELL/) PELL R.
XX
PI Reddy SA, Larrinua MI, Ruegger M, Weglarz T, Blakeslee B;
PI Oriedo VBJ, Savickas JP, McCrery AD, Miller Ab, Pogue PG;
PI Della-Cioppa RG, Wolfe MG, Zheng W, Gachotte D, Grosley R, Pell R;
XX
DR WPI; 2002-164823/21.
XX
PT Polynucleotide and amino acid sequences identified in one or more
PT metabolic pathways that lead to dwarfism and stunning in plants, useful
PT in agriculture to create dwarf varieties of any plant species -

Disclosure; Fig 9; 717pp; English.

CC The invention relates to polynucleotide and amino acid sequences
CC identified in one or more metabolic pathways that lead to dwarfism and
CC stunning in plants. Also included are vectors comprising the
CC polynucleotides, transgenic plants (including the seed and leaf)
CC transfected with the polynucleotides or vectors, a process for altering
CC the metabolism of a plant comprising providing the above vector and a
CC plant, and transfecting the plant with the vector under conditions such
CC that the metabolism of the plant is altered by expression of the isolated
CC nucleic acid from the vector, e.g. such that a stunting phenotype in an
CC industrial plant is produced and a process for the characterisation of
CC fractionated biological samples, comprising (a) providing one or more
CC fractionated biological samples, references samples, a gas chromatography
CC apparatus, a mass spectroscopy apparatus or data analysis software and
CC (b) treating the fractionated biological samples and the reference
CC samples with the gas chromatography apparatus to generate chromatographic
CC data corresponding to the fractionated biological samples and the
CC reference samples, (c) treating the fractionated biological samples and
CC the reference samples with the mass spectroscopy apparatus to generate
CC spectroscopic data corresponding to the fractionated biological samples
CC and the reference samples and (d) processing the chromatographic and the
CC spectroscopic data with the data analysis software. The nucleic acid and
CC the vector are useful for altering the metabolism of a plant and for
CC stunting a plant. The nucleic acids are useful in agriculture to create
CC dwarf varieties of any plant species. The present sequence is a
CC plant cDNA contig or singleton (related to dwarfism/stunting) identified
CC by searching a nucleic acid database with plant EST (expressed
CC sequence tag) and a BLAST (basic local alignment tool) stringency
CC of e-20.
XX
SQ Sequence 663 BP; 150 A; 204 C; 150 G; 159 T; 0 other;

Query Match   3.0%; Score 44.8; DB 24; Length 663;
Best Local Similarity 50.8%; Pred. No. 0.0023;
Matches 134; Conservative 0; Mismatches 127; Indels 3; Gaps 1;

```
Qy   316 CACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCGTCATGGCTCAC 375
         | |||  | ||  |||||  |||| ||||| |||  |  || ||||||| || || ||||
Db   250 CTCCTCCCTCAAATGGGTTGGGCTCCGACCCAACCCTACCTGCCTTTCCTCCTCCAACTCCCAC 309
```

```
Qy  376 TTTGANTGTGGTGCCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCCATTC---TCTGGGTTG 432
            ||||| || ||         | | |  |         |  |     ||       ||
Db  310 ACCGANTGTCGTTCCTTCGAGGACTTCCAAGTGGACTCACCATCCGGTCCAAGCTCACATG 369

Qy  433 GAGCCTGATGGCGTCATCTATGGTGGGGGCCACACTGGACGACAAGAACTCTGTGATGGCA 492
            ||| || ||||||| |||||||   | |  ||    | |||  |   |    || |
Db  370 GACCACCATGGCGACATCTATGCCAGGGGTTCCAGGACAATGAAGTGCGTCGGGATGCAG  429

Qy  493 TTACTGCAGGCCTTGGAGCTTCCTGCTGATCAGGAAGTACATCCCCGAAGAATCTTTCTTC 552
            | || |||||| |   ||| | |  | | |  |  ||| ||  ||| |  ||| || |
Db  430 TACCTCGAGGCCATACGCAAGCTCCAGGCTTCTGGCTTCAAGCCACTCCGATCCGGTCTAT 489

Qy  553 ATTTCTCTGGGCCATGATGAGGAG 576
            | ||  | ||||| ||||| |||
Db  490 CTCTCCTTCGTCCCCGATGAAGAG 513
```

OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run On: July 1, 2003, 04:22:44 ; Search time 2340 Seconds
                    (without alignments)
                    10444.013 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1 atggctcaagcggtgcgtttg..........ctcacctgcacaaactgtga 1509
Scoring table: IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                    Maximum Match 100%
                    Listing first 45 summaries

Database :

```
 1: em_estba:*
 2: em_esthum:*
 3: em_estin:*
 4: em_estmu:*
 5: em_estov:*
 6: em_estpl:*
 7: em_estro:*
 8: em_htc:*
 9: em_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	737.8	48.9	741	12	BG697810	BG697810 602661011
2	726.4	48.1	773	12	BG740058	BG740058 602631131
3	724.2	48.0	805	12	BG741998	BG741998 602633443
4	695.8	46.1	769	12	BG742908	BG742908 602632491
5	693.2	45.9	970	12	BG674119	BG674119 602619926
6	690.2	45.7	771	12	BG675035	BG675035 602621268

7	532.6	35.3	1023	12	BG743823	BG743823	602632667				
8	529.4	35.1	847	13	BI143597	BI143597	602907430				
9	528.4	35.0	575	13	BI832659	BI832659	603082042				
10	518.4	34.4	853	13	BI218153	BI218153	602934270				
11	465.2	30.8	706	13	BI101355	BI101355	602887103				
12	459.2	30.4	851	12	BF234135	BF234135	602024884				
13	424.2	28.1	748	12	BF234647	BF234647	602028464				
14	421.6	27.9	666	10	BB613230	BB613230	BB613230				
15	410.2	27.2	580	12	BG608571	BG608571	307199 MA				
16	406.2	26.9	580	10	BE032236	BE032236	131103 MA				
17	404.2	26.8	614	12	BF525111	BF525111	UI-R-AC0-				
18	401.6	26.6	544	12	BG608554	BG608554	307177 MA				
19	399.4	26.5	637	10	BI619403	BI619403	BB619403				
20	396	26.2	612	13	BI219686	BI219686	602935391				
21	394.4	26.1	565	10	BE032261	BE032261	131134 MA				
22	385.6	25.6	561	12	BM088240	BM088240	501660 MA				
23	385.6	25.6	584	13	BF079682	BF079682	230350 MA				
24	377.6	25.0	576	13	BI339966	BI339966	365090 MA				
25	377.6	25.0	576	13	BI339967	BI339967	365091 MA				
26	350	23.2	849	12	BF533743	BF533743	602073953				
27	340	22.5	717	13	BI147849	BI147849	602912738				
28	313	20.7	822	13	BI148134	BI148134	602912305				
29	308.6	20.5	567	12	BG895222	BG895222	358510 MA				
30	277.2	18.4	1267	12	BG674056	BG674056	602620050				
31	252.2	16.7	338	10	AW663138	AW663138	hh75b07.y				
32	250.4	16.6	711	14	BQ603860	BQ603860	MI-P-CP1				
33	247	16.4	430	10	BB845588	BB845588	BB845588				
34	245.4	16.3	385	10	BB872723	BB872723	BB872723				
35	239.2	15.9	742	13	BI247973	BI247973	602958833				
36	224.6	14.9	637	13	BM440197	BM440197	pgrin.pk0				
37	223.2	14.8	717	9	AJ456300	AJ456300	AJ456300				
38	220	14.6	368	10	BB871294	BB871294	BB871294				
39	195	12.9	378	10	BB872712	BB872712	BB872712				
40	194.8	12.9	350	10	BB869217	BB869217	BB869217				
41	192.8	12.8	321	9	AI317512	AI317512	uj23f05.y				
42	192.6	12.8	416	9	AI317514	AI317514	uj23f07.y				
43	188.0	12.5	533	14	BQ074284	BQ074284	fz27a06.y				
44	186.6	12.4	490	17	BH044325	BH044325	RPCI-24-2				
45	177.4	11.8	544	13	BI672970	BI672970	ft40b08.y				

ALIGNMENTS

RESULT 1
BG697810
LOCUS   BG697810   741 bp   mRNA   linear   EST 07-MAY-2001
DEFINITION   602661011F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4804202 5',
                mRNA sequence.
ACCESSION   BG697810
VERSION   BG697810.1 GI:13964441
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 741)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
                Email: cgapbs-r@mail.nih.gov
                Tissue Procurement: James Cleaver, M.D.
                cDNA Library Preparation: Life Technologies, Inc.
                cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
                Sequencing by: Incyte Genomics, Inc.
                Clone distribution: MGC clone distribution information can be
                found through the I.M.A.G.E. Consortium/LLNL at:
                http://image.llnl.gov
                Plate: LLAM10701 row: e   column: 03
                High quality sequence start: 4
                High quality sequence stop: 732.
FEATURES   Location/Qualifiers

RESULT 2
BG740058

LOCUS   BG740058   773 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION 602631131F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4776390 5',
            mRNA sequence.
ACCESSION   BG740058
VERSION   BG740058.1 GI:14050711
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 773)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10628 row: n column: 07
            High quality sequence stop: 773.
FEATURES   Location/Qualifiers
     source   1..773
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4776390"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   172 a   204 c   218 g   179 t
ORIGIN

Query Match
Best Local Similarity   99.8%;
Matches 727; Conservative 0; Pred. No. 4.3e-188; Mismatches 1; Indels 0; Gaps 0;
                                       Score 726.4;   DB 12;   Length 773;

Qy   1 ATGGCTCAGCCGGTGCCGTTTGCCGTTCGCTCGGCCCTCGGTGGCTCTATGGCTCCTCCCCAGTTTTTCCCT   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   46 ATGGCTCAGCCGGTGCCGTTTGCCGTTCGCTCGGCCCTCGGTGGCTCTATGGCTCCTCCCCAGTTTTTCCCT 105

Qy   61 ACCGTCTCCAGATCGATGGGCCCCGAGAGGGGCCATCAAAGGGCGTCGCGGAATCCCT 120
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 106 ACCGTCTCCAGATCGATGGGCCCCGAGAGGGGCCATCAAAGGGCGTCGCGGAATCCCT 165

Qy 121 TCTCAGTTCAGAAAAGAGGAAACGGCGTGGCGATGAAAGAGGGGGTGCCATCCAG 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 166 TCTCAGTTCAGAAAAGAGGAAACGGCGTGGCGATGAAAGAGGGGGTGCCATCCAG 225

Qy 181 ATTCCAACAGTGACTTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGAGTTCGGA 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 226 ATTCCAACAGTGACTTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGAGTTCGGA 285

Qy 241 AAATACAATTCAAAAGTCTTTCCCTACACAGTGGTCACGACCAGCCAGTGAAGTC 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 286 AAATACAATTCAAAAGTCTTTCCCTACACAGTGGTCACGACCAGCCAGTGAAGTC 345

Qy 301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAAGGCTCGGACGCCCTAC 360
        |||||||||||||||||||||||||||||||||||||||||||||||||
Db 346 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAAGGCTCGGACGCCCTAC 405

Qy 361 CTGCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGGCTGGGAGGTG 420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 406 CTGCTGATGGCTCACTTTGATGTGGTGCCTGCCCCTGAAGAAGGCTGGGAGGTG 465

Qy 421 TTCTCTGGGTTGGAGCCGTGATGGCCGTCATCTATGGTCGGGGCACACTGGACGACAAGAAC 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 466 TTCTCTGGGTTGGAGCCGTGATGGCCGTCATCTATGGTCGGGGCACACTGGACGACAAGAAC 525

BASE COUNT   179 a   195 c   188 g   179 t
ORIGIN

     source   1..741
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4804202"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

Query Match
Best Local Similarity   99.7%;
Matches 739; Conservative 0; Pred. No. 3.1e-191; Mismatches 2; Indels 0; Gaps 0;
                                       Score 737.8;   DB 12;   Length 741;

Qy 272 TCAGCACCAGCTTTATCCAGCCATGGAAGTCGTTGGGAAGAGAGTATAGCCCACCGTTCACTATCC 331
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 TCAGCACCAGCTTTATCCAGCCATGGAAGTCGTTGGGAAGAGAGTATAGCCCACCGTTCACTATCC   60

Qy 332 AAGGCTCGGACACCCCAGCTTGCAGCCCTACCCTGATGGCTCACTTTGATGTGGTGCCTG 391
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 AAGGCTCGGACACCCCAGCTTGCAGCCCTACCCTGATGGCTCACTTTGATGTGGTGCCTG 120

Qy 392 CCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTGGACGGTGATGGGCCTCATCT 451
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 CCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTGGACGGTGATGGGCCTCATCT 180

Qy 452 ATGGTCGGGGCACACTGGACGACAAGAACACTGTGATGATTCATTCATTCCTAACTTCA 511
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 ATGGTCGGGGCACACTGGACGACAAGAACACTGTGATGATTCATTCATTCCTAACTTCA 240

Qy 512 TCCTGCTGATCAGGAAGTACATCCCCGAGAGATCTTTTCATTTTCTCTGGGCCCATGATG 571
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 TCCTGCTGATCAGGAAGTACATCCCCGAGAGATCTTTTCATTTTCTCTGGGCCCATGATG 300

Qy 572 AGGAGTCATCAGGGACAGCAGGGGCTCAGGAGGATCTCAGCGTACAGTCAGTCAAGGGCGTCC 631
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 301 AGGAGTCATCAGGGACAGCAGGGGCTCAGGAGGATCTCAGCGTACAGTCAGTCAAGGGCGTCC 360

Qy 632 ACCTAGCCTTCATTGTGGACAGGGGGGTCTTCATCTTGGATGATGATTTCATTCCTAACTTCA 691
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 361 AGCTAGCCTTCATTGTGGACAGGGGGGTCTTCATCTTGGATGATGATTTCATTCCTAACTTCA 420

Qy 692 AGAAGCCCATCGCCCTTGAGTCTCAGAGAAGGGTTCCATGAACCTCATGTCTGCAAG 751
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 421 AGAAGCCCATCGCCCTTGAGTCTCAGAGAAGGGTTCCATGAACCTCATGTCTGCAAG 480

Qy 752 TAAACATGAGACTTCAGGCCCACTCTTTCAGCTCTCCAAAGGAGACAAGCAAGCATTGGCATCCTTG 811
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 481 TAAACATGAGACTTCAGGCCCACTCTTTCAGCTCTCCAAAGGAGACAAGCAAGCATTGGCATCCTTG 540

Qy 812 CAGCTGCTTGTCAGCCGATTGGAGCAGAGACACACCAATGCCTATCATAATTGGAAGCGGGACAG 871
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 541 CAGCTGCTTGTCAGCCGATTGGAGCAGAGACACACCAATGCCTATCATAATTGGAAGCGGGACAG 600

Qy 872 TGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCCTTCCCTGTCAATATAATAATCCTGA 931
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 601 TGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCCTTCCCTGTCAATATAATAATCCTGA 660

Qy 932 GCAAACCCATGGCTCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCA 991
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 661 GCAAACCCATGGCTCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCA 720

Qy 992 ATGCAATAAATCAGGACCACCA 1012
        ||||||||||||||||||||||
Db 721 ATGCAATAAATCAGGAGACACCA 741

```
Qy    481 TCTGTGATGGCATTACTCGAGGCCTTGGAGCTCCTGCTCATCAGGAAGTACATCCCCCGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 TCTGTGATGGCATTACTCGAGGCCTTGGAGCTCCTGCTCATCAGGAAGTACATCCCCCGA 585

Qy    541 AGATCTTCTTCATTTCTCTGGGGCCATGATGAGTGAGGAGCATCAGGAAGAGGGCTCAGAGG 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 AGATCTTCTTCATTTCTCTGGGGCCATGATGAGTGAGGAGCATCAGGAACACAGGGGCTCAGAGG 645

Qy    601 ATCTCAGCCCTGCTACAGTCAAGGGGGTCCAGCTAGCCTTCATTGTGTGGACGAGGGGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    646 ATCTCAGCCCTGCTACAGTCAAGGGGGTCCAGCTAGCCTTCATTGTGTGGACGAGGGGGC 705

Qy    661 TTCAACTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCCTTGCAGCCATCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    706 TTCAACTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCCTTGCAGCCATCTCA 765

Qy    721 GAGAAGGG 728
          ||||||||
Db    766 GAGAAGGG 773
```

```
RESULT 3
BG741998
LOCUS        BG741998          805 bp    mRNA        linear    EST 15-MAY-2001
DEFINITION   602633443F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4778507 5',
             mRNA sequence.
ACCESSION    BG741998
VERSION      BG741998.1  GI:14052651
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 805)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
             Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10634 row: f  column: 12
             High quality sequence stop: 771.
FEATURES             Location/Qualifiers
     source          1..805
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4778507"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT     184 a   211 c   225 g   185 t
ORIGIN

Query Match     48.0%;  Score 724.2; DB 12;  Length 805;
Best Local Similarity   98.2%;  Pred. No. 1.8e-187;
Matches   743;  Conservative   0;  Mismatches  13;  Indels   1; Gaps   1;

Qy      1 ARGGCTCAGGCGGTGCGTTTGCGTGCTGGCCCCTGGCGTCCGCTATGCTGTGCTCCTACTTTTTCCCT 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     46 ATGGCTCAGGCGGTGCGTTTGCGTGCTGGCCCCTGGCGTGGCCATCTAGCTGTGCTCCTACTTTTTCCCT 105

Qy     61 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGGGGCCATCAAAAGGCGGTCCCGGAATCCCT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    106 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGGGGGAGTGTATCAAAAGGGGCTCGCGGAATCCCT 165
```

```
Db    121 TCTCAGTTCAGCAAAGAGGAAGACGCGTGCGCGGATGAAGAGGGCGGCTGGAAAGAGGTGGCCATCCAG 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    166 TCTCAGTTCAGCAAAGAGGAAGACGGGTCGGGATGAAGAGGGCACTGAAGGTGCCATCCAG 225

Qy    181 ATTCCAAACAGTGACTTTTACCTCTGAGAAGAGTCCAATACTACAGGCCCTGCTGAGTTCGGA 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    226 ATTCCAAACAGTGACTTTTAGCTCTGAGAAGAGTCCAATACTACAGCCCCTGCTGAGTTCGGA 285

Qy    241 AAAATACATTCCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTTATCCAGCAATGAAGTC 300
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    286 AAAATACATTCGTAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCAATGAAGTC 345

Qy    301 GTGGAAGAGTATAGGCCACACCTGTTCACTAATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346 GTGGAAGAGTATAGGCCACACCTGTTCACTAATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 405

Qy    361 CTGCTGATGGCTCACTTTGATGTGGTGCTGCCCCCTGAGAAGGGCACACTGGACGACAAGAAC 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    406 CTGCTGATGGCTCACTTTGATGTGGTGCTGCCCCCTGAGAAGGGCACACTGGACGACAAGAAC 465

Qy    421 TTCTCTCTGGGTTGGACGCGTGATGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAAC 480
          |||||||||||||||||||||||||||||||||||||||
Db    466 TTCTCTGGGTTGGACGCGTGATGGCGTCATCTATGGTCGGGGGCCACACTGGACGACAAGAAC 525

Qy    481 TCTGTGATGGCATTACTCGAGGCCTTGGAGCTCCTGCTCATCAGGAAGTACATCCCCCGA 540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    526 TCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGCTCATCAGGAAGTACATCCCCCGA 585

Qy    541 AGATCTTCTTCATTTCTCTGGGGCCATGATGAGTGAGGAGCATCAGGAAGAGGGCTCAGAGG 600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    586 AGATCTTCTTCATTTCTCTGGGGCCATGATGAGTGAGGAGCATCAGGAACACAGGGGCTCAGAGG 645

Qy    601 ATCTCAGCCCTGCTACAGTCAAGGGGGTCCAGCTAGCCTTCATTGTGTGGACGAGGGGGC 660
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    646 ATCTCAGCCCCTGGCTACAGTCAAGGGCGTCAGCTAAGCCTTCATTGTGGACGAGGGGGC 705

Qy    661 TTCAACTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCCTTGACTAGCAGTCTCA 720
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    706 TTCATCTTGGATGATTGCATTCCTAACTTCAAGAACGCCCATCGCCCTTGACTAGCAGTCTCA 765

Qy    721 GAGAGAGGGTTCCATGAACCTCATGCTGCTCAAGTAAACA 757
          |||||||||| ||||||||||||||||||||||||||||
Db    766 GAGAAGGG-TACATGAACCCTCATGCTGACAGTAAACA 801
```

```
RESULT 4
BG742908
LOCUS        BG742908          769 bp    mRNA        linear    EST 15-MAY-2001
DEFINITION   602632491F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4777726 5',
             mRNA sequence.
ACCESSION    BG742908
VERSION      BG742908.1  GI:14053561
KEYWORDS     EST.
SOURCE       human.
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 769)
  AUTHORS    NIH-MGC http://mgc.nci.nih.gov/.
  TITLE      National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished (1999)
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
             Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10632 row: e  column: 23
```

RESULT 5
BG674119
LOCUS   BG674119   970 bp   mRNA   linear   EST 01-MAY-2001
DEFINITION   602619926f1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4745407 5',
             mRNA sequence.
ACCESSION   BG674119
VERSION   BG674119.1   GI:13905515
KEYWORDS   EST.
SOURCE   human.
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 970)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/,
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: James Cleaver, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM10592   row: c   column: 08
             High quality sequence stop: 695.
FEATURES   Location/Qualifiers
     source   1..970
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4745407"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Skn3"
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   269 a   254 c   228 g   219 t
ORIGIN

Query Match   45.9%;   Score 693.2;   DB 12;   Length 970;
Best Local Similarity   91.2%;   Pred. No. 6.6e-179;
Matches 872;   Conservative 0;   Mismatches 63;   Indels 21;   Gaps 12;

[sequence alignment block — right column, Qy/Db pairs]

ACCESSION section (left column):

High quality sequence stop: 767.
FEATURES   Location/Qualifiers
     source   1..769
                     /organism="Homo saplens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4777726"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   211 a   221 c   169 g   168 t
ORIGIN

Query Match   46.1%;   Score 695.8;   DB 12;   Length 769;
Best Local Similarity   99.7%;   Pred. No. 1.1e-179;
Matches 697;   Conservative 0;   Mismatches 2;   Indels 0;   Gaps 0;

RESULT 6

BG675035

LOCUS   BG675035   771 bp   mRNA   linear   EST 01-MAY-2001
DEFINITION   602621268F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4746689 5',
               mRNA sequence.
ACCESSION   BG675035
VERSION   BG675035.1   GI:13906431
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 771)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
               Email: cgapbs-r@mail.nih.gov
               Tissue Procurement: James Cleaver, M.D.
               cDNA Library Preparation: Life Technologies, Inc.
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
               Clone distribution: MGC clone distribution information can be
               found through the I.M.A.G.E. Consortium/LLNL at:
               http://image.llnl.gov
               Plate: LLAM10595 row: h column: 18
               High quality sequence stop: 750.
FEATURES   Location/Qualifiers
     source   1..771
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4746689"
                     /clone_lib="NCI_CGAP_Skn3"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   170 a   203 c   220 g   178 t
ORIGIN

Query Match   45.7%; Score 690.2; DB 12; Length 771;
Best Local Similarity 98.1%; Pred. No. 3.8e-178;
Matches 709;   Conservative 0;   Mismatches 13;   Indels 1;   Gaps 1;

Qy   1 ATGGCTCAGCGCTGCGTTTGCGTGCGTCGTGGCGCCTGGTGGCTATGCGTGCTCCTAGTTTTTCCT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   46 ATGGCTCAGCGCTGCGTTTGCGTGCGTCGTGGCGCCTGGTGGCTATGCGTGCTCCTAGTTTTTCCT 105

Qy   61 ACCGTCTCCAGATCGATGGGCCCGAGCAGGCGGGGAGCAGGAGCATCAAAAGGGCCGTGCGAATCCCT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   106 ACCGTCTCCAGATCGATGGGCCCGAGCAGGCGGGGAGCAGGAGCATCAAAAGGGCCGTGCGAATCCCT 165

Qy   121 TCTCAGTTCAGCAAAGAGGAAGCCGTGCGGATGAAAGAGGCTGAAAGTGCCATCCAG 180
           |||||||||||||||||||||||||||||||||||||||||||||| ||||||||||||
Db   166 TCTCAGTTCAGCAAAGAGGAAGCCGTGCGGATGAAAGAGGCTGAAAGTGCCATCCAG 225

Qy   181 ATTCCAACAGTGACTTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCCTGAGTTCGGA 240
           |||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   226 ATTCCAACAGTGACTTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCCTGAGTTCGGA 285

Qy   241 AAATACATTCATAAAAGTCTTTCCTTACAGTGTGTCAGCACCACATACGACCATGAAGTC 300
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   286 AAATACATTCGTAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTC 345

Qy   301 GTGGAAGAGTATAGCCACCTGTTCACTATTCCAA-GGCTCGGACCCCAGAGTTGCCAGCCCTA 359
           ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||||
Db   346 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGGGCTCGGACCCCAGCTTGCAGCCCTA 405

Qy   360 CCTGCTGATGGCTCACTTTGATGTGGTAGTCCCTGCCCCTGAAGAGGCTGGGGGAGGTGCCCC 419
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   406 CCTGCTGATGGCTCACTTTGATGTGGTGGTCCCTGCCCCTGAAGAGGCTGGGGGAGGTGCCCC 465

Qy   420 ATTCTCTCTGGGTTGGGAGCGTGATGGGGCGGCGGTCGGGCGGCCACACTGGGACGACAAGAA 479
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   466 ATTCTCTCTGGGTTGGGAGCGTGATGGGGCGGCGGTCGGGCGGCCACACTGGGACGACAAGAA 525

Qy   480 CTCTGCGATGGCATTACTGCAGCGGCCCTTGGAGCTCAGCTAGCCCTTCATTGTGGGACGAGGGG 539
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   526 CTCTGCGATGGCATTACTGCAGCGGCCCTTGGAGCTCAGCTAGCCCTTCATTGTGGGACGAGGGG 585

Qy   540 AAGATCTTTACTTCATTTCTCTGGGCCCATGAGGAGGTCATCAGGGACGAGGGGCTCAGAG 599
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   586 AAGATCTTTACTTCATTTCTCTGGGCCCATGAGGAGGTCATCAGGGACAGGGGGCTCAGAG 645

Qy   600 GATCTCAGCCCTGCTACAGTCAAGGGGCGGGTTCAGCTAGCCTTCATTGTGGACGAGGAG 659
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   646 GATCTCAGCCCTGCTACAGTCAAGGGGCGGTCCAGCTAGCCTTCATTGTGGACGAGGAG 705

Qy   660 CTTCATCTTGGATGAGATTTCATTCCTAACTTCAAGAAGCCCATGCGCCTTGATTGCAGTGTC 719
           |||||||||||||||||||||||||||||| ||||||||||||| ||
Db   706 CTTCATCTTGGATGAGATTTCATTCCTAACTTCAAGAAGCCCATGCGCCTGTATGCGGTCTCGAG 765

Qy   720 AGA 722
           |||
Db   766 AGA 768

RESULT 7

BG743823

LOCUS   BG743823   1023 bp   mRNA   linear   EST 15-MAY-2001
DEFINITION   602632667F1 NCI_CGAP_Skn3 Homo sapiens cDNA clone IMAGE:4777745 5',
               mRNA sequence.
ACCESSION   BG743823
VERSION   BG743823.1   GI:14054476
KEYWORDS   EST.
SOURCE   human.
  ORGANISM   Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 1023)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)

```
LOCUS       BI143597                847 bp    mRNA    linear    EST 05-JUL-2001
RESULT 8
BI143597
```

```
DEFINITION  6029074305F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5064342
            5', mRNA sequence.
ACCESSION   BI143597
VERSION     BI143597.1  GI:14603598
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 847)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11173 row: l column: 07
            High quality sequence stop: 806.
FEATURES             Location/Qualifiers
     source          1..847
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:5064342"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT        193 a       239 c       222 g      193 t
ORIGIN
```

```
Query Match         35.1%;  Score 529.4;  DB 13;  Length 847;
Best Local Similarity  81.5%;  Pred. No. 5.6e-134;
Matches  674;  Conservative  0;  Mismatches  146;  Indels  7;  Gaps  5;

Qy     1  ATGGCTCAGCGGGTGGCGGTTGGCGTGGCGCTATGCGTCCTCCAGTTTTT-CCC   59
          |||| |  ||  | |||| |||    | |       |                 |||
Db     3  ATGGCTGAGCTACTTGCTAGCTTTGCCCCCGTCCCCGGTGGCTCCTTCTCTATTTTTACGC   62

Qy    60  TACCGTCTCCAGATCGATGGGCCCCGGAGGACGCGGGAGCATCAAAGGCGTCCGGAATCCC  119
          |||  ||||||||  || ||||  || |  |||  ||||||  |  |||  |||||||||||
Db    63  TACGGTCTCCGGATCCCACTGGGCCCTAGAAGCACGGGAAAATGGGGGGCGTCCGGATCCC  122

Qy   120  TTCTCAGTTGCAGCAAAGAGGAAGCGCGTGGCCGATGAAGAAGAGGCGCCTGAAAGGTTCG  179
          |||  |||  | ||||  | | |||  |  || |  | || || |  || || ||  |||||
Db   123  TTCCCAGTTCAGCGAGGAGGAGGCGGTGTGCTATAAAAGAGGGCTGAAAGGTGCCATCCA  182

Qy   180  GATTCCAACAGTGACTTTTAGCTGTGACAAGTCCAAATACTACAGCGCCCTGGCTGAGTTGG  239
          |||    | ||| | |   || | ||    |  ||  || || | ||  ||| |  |||||
Db   183  GATTCCCACAGTGTNCTTCAGCCGACGAGGAATCCAACAACCAGCCCTTGCTGACGAGTTCG  242

Qy   240  AAAATACATTCATAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGT  299
          || || |  ||   |||  | ||| |||||  | |||||||  || || || | |||||
Db   243  AGAATATACCGACAAAGCCTTCCCTACAGTGTTCCACAGCCCTTGTCCAACATGAAGT   302

Qy   300  CGTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCCTCGGACCCCAGCTTGCAGCCTA  359
          | |  |||| |||||  |  || ||| || |  |||| ||| ||| |||| || ||||||
Db   303  CGTGGCAAAGTATAGCACCACCTGTTCCACCATCCAAGGCTCAGACCCCAGTTTGCAGCCCTA  362

Qy   360  CCTGCTGAATGGCTCACTTTGATGTGGTGGTGGGCCTGAAGAAGAGGTCGGGAGGTGCCCC  419
          ||||||||  |||| ||  ||||| ||||| || ||| ||||| |   ||| ||||| ||
Db   363  CATGCTGATGGCTCACATTGATGTGGTTGTTCCTGCCCCGGAAGAAGAGAAGTGCCCCC  422

Qy   420  ATTCTCTGGGTTGAGCGTGAATGGCGTCATCTANGGGTCGGCGCACACTGGACGACAAGAA  479
          ||||||| || |   |  |  |||| |   | |    |||  | | || ||| | || |||
Db   420  ATTCTCTGGGTTGAGCGTGAATGGCGTCATCTANGGGTCGGCGCACACTGGACGACAAGAA  479
```

```
LOCUS
RESULT 8
BI143597
```

```
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: James Cleaver, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA
            Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10632 row: f  column: 18
            High quality sequence stop: 749.
FEATURES             Location/Qualifiers
     source          1..1023
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:4777745"
                     /clone_lib="NCI_CGAP_Skn3"
                     /note="Organ: skin; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.5kb. Library constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT        281 a       279 c       223 g      240 t
ORIGIN
```

```
Query Match         35.3%;  Score 532.6;  DB 12;  Length 1023;
Best Local Similarity  99.1%;  Pred. No. 8.4e-135;
Matches  546;  Conservative  0;  Mismatches  4;  Indels  1;  Gaps  1;

Qy   960  AAGCAGTTTATTGGAGAGAGAAAATCCCTTAACCAAATGCAAATAATCAGGACCACACGGCACT 1019
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    14  AATGAGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAAATAATCAGGACCACACGGCACT   73

Qy  1020  CACCATAATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCACACCGGCACAGT 1079
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    74  CACCATAATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCCCACACAGT  133

Qy  1080  CAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAAGAGGTCCTAGAACTCACGAAGAACAT 1139
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   134  CAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAAGAGGTCCTAGAACTCACGAAGAACAC  193

Qy  1140  TGTGGCTGANTAACAGAGTCCAGTTCCATTGTGTTGAGTGGCCTTTGACCCCTTCCCGTCAG 1199
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   194  TGTGGCTGANTAACAGAGTCCAGTTCCATTGTGTTGAGTGGCCTTTGACCCCTTCCCGTCAG  253

Qy  1200  CCCTTCTCGATGACAAGGCCTTGGGCTACCAGCGTGCTCCGCCAGACCGTACACAGTCCGTCTT 1259
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   254  CCCTTCTCGATGACAAGGCCTTGGGCTACCAGCGTGCTCCGCCAGACCGTACACAGTCCGTCTT  313

Qy  1260  CCCGGAAGTCAATATTACTG-CCCCAGTGTACTTCTATTGGCAACACAGCAGCAGCCATTCT 1318
          |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db   314  CCCGGAAGTCAATATTACTGCCCCCAGTTACTTCTATTGGCAACACAGCAGCAGCCATTCT  373

Qy  1319  TTACAAAGCTCACCACTGGCANTCTACCCCATCTACACTACAGCCTGAAGACT 1378
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   374  TTACAAACCTCACCACTGGCANTCTACCCCATCTACACTACAGCCTGAAGACT  433

Qy  1379  TCAAACGCCATCCATGGAGTCAACGAGAAAACTCAGTCCAACCCTATGAGACCCAAGTGA 1438
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   434  TCAAACGGCATCCATGGAGTCAACGAGAAAATCTCAGTCCATGAGACCCAAGTGA  493

Qy  1439  AATTCATCTTTGAGTTGATTCAGAAATGCTGACACAGACCAGAGCCAGTTTCTCACCTGC 1498
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   494  AATTCATCTTTGAGTTGATTCAGAAATGCTGACACAGACCAGAGCCAGTTTCTCACCTGC  553

Qy  1499  ACAAACTGTGA 1509
          |||||||||||
Db   554  ACAAACTGTGA  564
```

```
Qy          1 ATGGCTCAGGGTGCGTGCGTTTGCGTGCCTGCGGCCGCCCTGGTGCCTGGGCCTATGCCTGCTCCTCCTAGTTTTCCCT   60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         46 ATGGCTCAGGGTGCGTGCGTTTGCGTGCCTGCGGCCGCCCTGGTGCCTGGGCCTATGCCTGCTCCTCCTAGTTTTCCCT  105

Qy         61 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGAGCAGCAGCATCAAAAGGGGTCGGCAAATCCCT  120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        106 ACCGTCTCCAGATCGATGGGCCCGAGGAGCGGGAGCAGCAGCATCAAAAGGGGTCGGCAAATCCCT  165

Qy        121 TCTCAGTTCAGCAAAAGAGGAAACGCGTCCGGATGAAAGAGGGAGCCATCCAG  180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        166 TCTCAGTTCAGCAAAAGAGGAAACGCGTCCGGATGAAAGAGGGAGCCATCCAG  225

Qy        181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGAGTTCGGA  240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        226 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGAGTTCGGA  285

Qy        241 AAATACATTCATTAAAGTTCTTTCCTCACAGTGGTTCAGCACCAGCTTTATCCAGCATGAAGTC  300
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        286 AAATACATTCATTAAAGTTCTTTCCTCACAGTGGTTCAGCACCAGCTTTATCCAGCATGAAGTC  345

Qy        301 GTGGAAGAGTATAGGCCACCTGTTCACTATCCAAGGCCCACCCCAGCTTCGACGCTTGCAGCCCTAC  360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        346 GTGGAAGAGTATAGGCCACCTGTTCACTATCCAAGGCCCACCCCAGCTTCGACGCTTGCAGCCCTAC  405

Qy        361 CTGCTGATGGCCTCACTTTTGATGTGTGTCCCCTGAGCAGAAGGCTGGGAGGTGCCCCA  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        406 CTGCTGATGGCCTCACTTTTGATGTGTGTCCCCTGAGCAGAAGGCTGGGAGGTGCCCCA  465

Qy        421 TTCTCTGGGGTTGGACGTGATGGGCGGTCATCTATGTCGGGGCACACTGACGACGACAAGAAC  480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        466 TTCTCTGGGGTTGGACGTGATGGGCGGTCATCTATGTCGGGGCACACTGACGACGACAAGAAC  525

Qy        481 TCTGTGGATGGCCATTACTGCAGGCCCTTGGAGCTCCTGGCTGATCAGGAAGTA  530
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        526 TCTGTGGATGGCCATTACTGCAGGCCCTTGGAGCTCCTGGCTGATCAGGAAGTA  575
```

RESULT 10
```
LOCUS       BI218153                 853 bp    mRNA    linear     EST 11-JUL-2001
DEFINITION  602934270F1 NCI_CGAP_Li9 Mus musculus cDNA clone IMAGE:5097737 5',
            mRNA sequence.
ACCESSION   BI218153
VERSION     BI218153.1  GI:14671597
KEYWORDS    EST.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 853)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: cgapbs-r@mail.nih.gov
            Email: cgapbs-r@mail.nih.gov
            Jeffrey E. Green, M.D.
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM11234 row: k column: 18
            High quality sequence stop: 835.
FEATURES             Location/Qualifiers
     source          1..853
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
```

```
Db        423 GTTCTCAGGCCTGGAACCCAATGGCTTCATCTATGGCCGGGGTGCCTGGGACAACAAAA  482

Qy        480 CTCTGTGATGGCCATTACTGCAGGCCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCCCG  539
              |||||||||||||||||||||||||||||||||||||||||||||||||||||| |  |||
Db        483 CTCTGTGATGGCCATCCTCCATGCTTTTGGAGCTCCTGTTGATCAGAAACTACACGCCCAA  542

Qy        540 AAGATCTTTCTTCATTTCTCTGGGCCATGAATGAGGAGTCATCAGGG---ACAGGGGGCTCA  596
              |||||||||||||||||||||                         |      ||||||||||
Db        543 AAGATCTTTCTTCATTTGTTTGGGCCATGGATGAGGAGTTGTGGGGAAAAGGGGGCTCA  602

Qy        597 GAGGATCTCAGCCCTGCTACAGTCAAGGGGGGTCCAGCTAGGCTGCCTTCATTGTGGACGAGGG  656
              | |||||||||||||||||||||||||||||||||||||||||||||||||||| || ||
Db        603 GAAGATCTCAGCCACTCTTACAGCCGAAGGGGGTCCAGCTAGGCTGCCTTCCTTGTGGATGAAGG  662

Qy        657 GGGCTTC-ATCTTGGATGATTCATTCCTTAAACTTCACAGAAGGCCCATCGCCTTGATTGCAG  715
              |||||||  ||||||||||||| |||||||||||||| |||||| ||||||||||||||||||
Db        663 GAGCTTCAATCTTGGAAGGCCTTCATTCCAAAACCTCGAGAAGCCAGTTGCCATGATTTTCAG  722

Qy        716 TCTCAGGAAGGGGTTCCATGAACCCTCATCGTCGCAAGTAAAACATGACTTCAGGCCCACTCTT  775
              |||||||||||||||||| |||||||||||||||||||||||| |||| ||||||| ||||||
Db        723 TCACTGAGAAGGGTTGCCCTCATCGTCATCGCAAGTCAACACATGACTCCAGG-CACTCTT  781

Qy        776 CAGCTCCTCCAAAGGAGAGACAAGCAATTGGCATCCTTGCAGCTGCTGTC  822
              |||||| ||| |||| ||| |||||||||||| ||| | |||| ||||||
Db        782 CAGCTCC-CCCAAGGAGACAAGCAATTGGCATTGCTTTCTTGCCGCTGTC  827
```

RESULT 9
```
LOCUS       BI832659                 575 bp    mRNA    linear     EST 04-OCT-2001
DEFINITION  6030042F1 NIH_MGC_120 Homo sapiens cDNA clone IMAGE:5221187 5',
            mRNA sequence.
ACCESSION   BI832659
VERSION     BI832659.1  GI:15944209
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 575)
AUTHORS     NIH-MGC http://mgc.ncl.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Life Technologies, Inc.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM1556 row: c column: 12
            High quality sequence stop: 575.
FEATURES             Location/Qualifiers
     source          1..575
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:5221187"
                     /clone_lib="NIH_MGC_120"
                     /lab_host="DH10B"
                     /note="Organ: pooled pancreas and spleen; Vector:
                     pCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
                     source anonymous pool of spleen and pancreas from 28 yo
                     male. Library is oligo-dT primed and directionally cloned
                     (EcoRV site is destroyed upon cloning). Average insert
                     size 1.5 kb, insert size range 1-2.5 kb. Library is
                     normalized and enriched for full-length clones and was
                     constructed by C. Gruber (Invitrogen). Research Genetics
                     tracking code 025. Note: this is a NIH_MGC Library."
BASE COUNT      128 a    152 c    165 g    130 t
ORIGIN
```

Query Match   529; Best Local Similarity 99.8%;
Matches 529; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Best Local Similarity   99.8%; Score 528.4; DB 13; Length 575;
Conservative   0; Pred. No. 8.2e-134;
Mismatches   1; Indels   0; Gaps   0;

RESULT 11
BI101355

LOCUS   BI101355   706 bp   mRNA   linear   EST 26-JUN-2001
DEFINITION 6028871O3F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:5042305
           5', mRNA sequence.
ACCESSION   BI101355
VERSION   BI101355.1 GI:14552248
KEYWORDS   EST.
SOURCE   house mouse.
  ORGANISM Mus musculus
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
          Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 706)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: Jeffrey E. Green, M.D.
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Incyte Genomics, Inc.
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LLAM1l116 row: f column: 02
          High quality sequence stop: 704.
FEATURES   Location/Qualifiers
    source   1..706
           /organism="Mus musculus"
           /strain="FVB/N"
           /db_xref="taxon:10090"
           /clone="IMAGE:5042305"
           /clone_lib="NCI_CGAP_Kid14"
           /lab_host="DH10B (T1 phage-resistant)"
           /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
           Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
           Average insert size 1.75 kb. Constructed by Life
           Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT   160 a   194 c   190 g   162 t
ORIGIN

Query Match   30.8%; Score 465.2; DB 13; Length 706;
Best Local Similarity 81.0%; Pred. No. 2.1e-116;
Matches 567; Conservative 0; Mismatches 128; Indels 5; Gaps 2;

Qy   1 ATGGCTCAGCGCGGTGCGGTTTGCGTTGCGTGCTCCTGGCCTATGCGTCGTCCCTAGTTTTCGCTTTCCCT 60
      | |||||| | | | |||||||| |||||||||
Db   7 ATGGCTGAGCTACTTGCTAGCTTTGCCCCCGTCCGCGGTGGCTAGTGCTCCTGCCCTTCGTTTTCTTTTCGCT 66

Qy   61 ACCGGTCTCCAGATCGATGGGCCCGGAGAGCGGGAAGGGCGGGAGCATCAAAAGGGCGGGAATCCCT 120
      || ||||||| || ||| | ||||||||||||||
Db   67 ACGGTCTCCGGATCCACTGGCCCTAGAAGCAGGGAAAATCGGGGGGCGGGTCCCGGGGATCCCT 126

Qy 121 TCTCAGTTCAGCAAAGAGGAGAACGCGTCCGATGAAAAGAGGCGCTGAAAGGTGCCCATCCAG 180
      || ||||||| ||| ||| ||||||||||||||
Db 127 TCCCAGTTCAGCGAGGAGGCGTGTCGGCTATAAAAGGCGGTGCCCATCCAG 186

Qy 181 ATTCCCAACAGTGACTTTTAGCTCTGAGAAGTCAATACTACAGCGCCTGGCGTTTATCCAGCATGAAGTC 240
      ||||||||| | ||||||||||||||||||
Db 187 ATTCCCACAGTGCTTTCAGCCACCAGAGAATCCAACACACAGCCCTTGCTGAGTTTGGA 246

Qy 241 AAATACAATTCATAAAGTCTTTCCTACAGTGGTCAGAACAGCCTTATCCAGCATGAAGTC 300
      ||||| | |||||||||||||||||||||||
Db 247 GAATATATCCGCAAAGCCTTCCCTACAGTGTTCACAGCAGTGGTCAGAACATGAAGTC 306

Qy 301 GTGGAAGAGTATAGCCACCTGTTCACTATTCCAAAGGCTCGGACCAGCCTTGCAGCCCTAC 360
      |||| |||||||||||||||||| |||||||||||||||||||
Db 307 GTGGCAAAGTATAGCCACCTGTTCACCATCCAAAGGCTTCAGACCCCAGCCTTGCAGCCCTAC 366

Qy 361 CTGCTGATGGCTCCACTTTTGATGTGGTGGCCTGCCCCCCTGAGAGAAGGGCTGGGAGGTGCCCCCA 420
      |||||||| |||| |||||||||||||||||||||||
Db 367 ATGCTGATGGCTCACATTGATGTGGTTCCTGCCCCCGGAAGAAGGATGGGAGGGAGGTGCCCCCA 420

Qy 421 TTCTCTCTGGGTTGGACGCTGGCGTCATCTATAGGTCGGGGCGCCACACTGGACGACAGAGAAC 480

BASE COUNT   198 a   237 c   224 g   194 t
ORIGIN

Query Match   34.4%; Score 518.4; DB 13; Length 853;
Best Local Similarity 82.0%; Pred. No. 5.9e-131;
Matches 658; Conservative 0; Mismatches 136; Indels 8; Gaps 5;

Qy   39 TATGCTGGCTCCTAGTTTTCCTACGTCTCCAGATCGATGGGCCCGGAGGAGCGGGGAGCA 98
      ||| |||||||||||||||||||||||||
Db   43 TGTGCTCCTTCTCTTTTTCGGTCGTCCCACCTCAGCCTAGAAGCAGGGAAA 102

Qy   99 TCAAAGGGGGCTCGGAATCCCTTCTCAGTTCAGCAAAGAGGAACGCGTCCGATGAAAGA 158
      |||||||||||||||||||||||||||||||
Db 103 TCGGGGGGCGTCCCGGATCCCTTCCCAGTTCAGCGAGGAGGCGTGTCGGCTATAAAGA 162

Qy 159 GGGGGCTGAAAGGTGCCATCCAAGAGTTGACTTTTAGCTCTGAGAAGTCCAATAC 218
      ||||||||||||||||| | || ||||||| ||||
Db 163 GGCCGGTGAAAGGTGCCATCCAGATTCCCACAGTGCTTTCAGCCACCAGAGAATCCAACAC 222

Qy 219 TACAGCCCTGGCTGAGTTTCGGAAAAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCAGCAC 278
      ||||||| || | ||||| |||||||||||
Db 223 CACAGCCCTTGCTGAGTTTGGAGAATATATCCGCAAAGCCTTCCCTACAGTGTTCCACAG 282

Qy 279 CAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCTGTTCACTATTCCAAAGGCTC 338
      ||||| ||||||||||||| | ||||||||||||
Db 283 CAGCCTTGTCCAAACATGAAGTCGTGGCAAAGTATAGCCACCTGTTCACCATCCAAAGGCTC 342

Qy 339 GGACCCAGCCTTGCAGCCCCTACCGCTGCTGATGGCTCCACTTTTGATGTGGTGGCCTGCCCCCTGA 398
      || || ||||| |||||||| ||||| |||||
Db 343 AGACCCCAGTTTGCAGCCCTACATGCTGATGGCTCACATTGATGTGGTTCCTGCCCCCGGA 402

Qy 399 AGAAGGCTGGGAGGTGCCGCCCCATTCTCTCGGGTTGGACGCTGCATCATCTATAGGTCG 458
      |||||| ||||||| || |||||| |||||
Db 403 AGAAGGATGGGAGGGAGGTGCCCCCCAGTTCTCACGGGTTCTTGGAGCTCATCTATAGGTCG 462

Qy 459 GGGGCACACTGGACGACAACGAACTCTGTGANGGCATTACTGCAGCCCTTGGAGCTCCTCGCT 518
      ||| ||||| |||||| || || ||||||
Db 463 GGGGTGCGCGTGGACAACAAAAACTCTGTGANGGCGATCCTGTTTGGGACGCTGGAGAGTC 522

Qy 519 GATCAGGAAGTACATCCCCCGAAGAGATCTTTTCTCTCTGGGCCCATGAATGATGAGGAGTC 578
      || || |||||| | ||||||||||
Db 523 GATCAGAAACTACAGCCCCAAAAGAATCTTTCTTCATTCGTTTGGGCCATGAATGAGGAGGT 582

Qy 579 ATCAGGGACA---GGGGGCTCAGAGGAATCTCAGCCCTGCTACAGTCAAGGGGCCTCCAGCT 635
      || ||| ||||| ||| | |||||||||||
Db 583 GTCCGGGGGAACAGGGGGCTCGAGAGATCTCAGCACTCTTGGATTGATTTCATTCCTAACTTCAAGA 642

Qy 636 AGCTTCATTGTGGACGAGGGGGCCTTCAACTCTTGGATGATTTCATTCCTAACTTCAAGA 695
      ||||||| || |||||||||| |||||
Db 643 AGCCTTCCTTGTGTGGATGAAGGGACTTTATCTTGGAAGGCTCATTCCAAAAACTCGAGAA 702

Qy 696 GCCCCATCCGCC--TTGATTGCGAGCTCTGAGCCATCGTCTCGCGTGCTGGAGAAGTGCCAT 754
      ||| || ||||||||||||||||
Db 703 GCCAGTTGGCCAATGATTTCAGTCACTGAAAGGGTGCCTTG--ACCTCANTGCTGCAAGTAA 761

Qy 755 AC--ATGACTTCAGGCCACTCTTCAGCTCCTCCAAAGAGAGCAAAGCATTGGGCAATCCTTGC 812
      || || ||||||||||||| |||||| ||||
Db 762 ACCATTGACTCAGGCCACTCTTCAAGTCCCCAAAGGAGAAAAGCATTGCAT-CTTTC 820

Qy 813 AGCTGCTGCTGTCAGCGCCGATTGGAG 834
      || |||||| ||| ||||||
Db 821 TGGCGCGTGCTGCAAGCGACTGGGGAG 842

```
Db   667 GGAGCTTTATCTTGGAAGGCCTTCATTCCAAACCTCGAGAA 706
             ||| ||| ||||||||| ||| |||  || ||
Qy   656 GGGGGCTTCATCTTGGATGATTTCATTCATTCTCAAGAA 695
Db   667 AGAAGATCTCAGCACTCTTACAGGGAAAGGGTGTCCAGTGAGCCTTCCTTCATTGTGGACGAGG 666
Qy   607 AGAAGATCTCAGCACTCTTACAGGGAAAGGGTGTCCAGTGAGCCTTCCTTGTTGTGGATGAAG 655
Db   547 AGATCTTTCTTCACTTGCTTGCCCATGAGAGGAGGTGTCCCGGGAAAAAGGGGGCTC 606
Qy   541 AGATCTTTCTTCA-TTTCTCTGGGCCCATGAGAGGAGGAGTCATCAGG------ACAGGGGCTC 595
Db   487 TCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGTTGATCAGAAAACTACACCCCAAA 546
Qy   481 TCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGTTGATCAGAAAACTACACCCCAA 540
Db   427 TTCTCAGGCCTGGAACGCAATGGCTTCATCTATGGCCGGGGGTGGCTGACACAAAAAC 486
```

```
Query Match    30.4%;  Score 459.2; DB 12; Length 851;
Best Local Similarity  78.3%;  Pred. No.1.1e-114;
Matches  617;  Conservative   0; Mismatches 158; Indels  13; Gaps   5;

Qy    1 ATGGCTCAGGCGGTGCGGTTGCGTGCGTGGCCCTGGTGGGCCTGGGCGCGGAGGAGGCGGGAGGAGCATCAAAAGGGCGTCGGGAATCCCT 60
        ||||| ||||| || ||| |||||||||  ||||||||||||
Db   37 AUGGCTCAGCTACTTGCTAGCTTGCTGCCCGCCGCCCTGGGCAGCTGTGCCTCCTTCTCTTCTTTTTCGCT 96

Qy   61 ACCGTCTCCAGATCGATCGATGGGCCCGAGGAGCGGAGCATCAAAAGGGCGTCGGGAATCCCT 120
        |||||| ||||  ||| |||  || |||||
Db      ...
```

```
FEATURES             Clone distribution: MGC clone distribution information can be
                     found through the I.M.A.G.E. Consortium/LLNL at:
                     http://image.llnl.gov
                     Plate: LLAM9448 row: o column: 15
                     High quality sequence stop: 649.

        source       1..748
                     /organism="Mus musculus"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4163942"
                     /clone_lib="NCI_CGAP_Li9"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /note="Organ: liver; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.9 kb.Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT     171 a       206 c       201 g       170 t
ORIGIN

Query Match                28.1%;   Score 424.2;  DB 12;  Length 748;
Best Local Similarity      77.5%;   Pred. No. 3.9e-105;
Matches 579;  Conservative  0;  Mismatches 158;  Indels  10;  Gaps  5;

Qy      1 ATGGCTCAGGCGGTGGCTTTCGTGCTGCTGGCCCTGGTGGCTCAGGCGGTGGCTTTCGTGCTGCTGGCCCTGGT 60
          ||||||| ||||| |||  |||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db      1 ATGGCTGAGTCTACTTGCTAGCTTGCCCCTGGCAGCGTGTGTCATCAAAGAGGCGTCGCGAATCCCT 60

Qy     61 ACCGTCTCCAGATCGATGGGCCCCGAGGAGACGACGGGGGCATCAAAGAGGCGTCGCGAATCCCT 120
          |||||||||||| | |||  | ||||| ||||||||||||||||||||| |||||||||||||
Db     61 ACGGTCTCCGGATCCACTGGCCCTAGAAGCAGGGAAAATCGGGGGGCGTCCCGGATCCCT 120

Qy    121 TCTCAGTTCAGCAAAGAGGAAACGGCGTCGCCGGATGAAAAGAGTGCCATCCAG 180
          |||||||||||||||||||||||||||||||||||||| | | ||||||||||
Db    121 TCCCAGTTCAGCGAGGAGGAGCGCGTGTCGCTATATAAAAGAGGCGTCCATCCAG 180

Qy    181 ATTCCAACAGTGACTTTTAGCTCTGAGAGTCCAATACTACAGCCCTGGCTGACTTCGGA 240
          ||||||| |||||||||||||| ||||||| || ||||||||||||||||||||||||||
Db    181 ATTCCCACAGTGTCTTTCAGCCACGAGGAGAATCCAACACCACCAGCCCTTGCTGAGTTTGGA 240

Qy    241 AAAATACATTCATAAAGTCTTTCTACAGTGGTCCAGCAGCTTTATCCCAGCATGAAGTC 300
          |||||||||| | ||||||||| || ||| ||||||||||||||||||||||| ||||
Db    241 GAATATATCCGCAAAGCCTTCCCTACAGTTCCACAGCAGCCTTGTCCAACATGAAGTC 300

Qy    301 GTGGAAGAGTATAGCCACCTTGTTCACTATACCAAGGCTCGGACCCCCAGTTGCAGCCCTAC 360
          |||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 GTGGCAAAGTATAGCCACCTGGTTCACCATCACCAAGGCTCAGAGCCCCAGTTGCAGCCCTAC 360

Qy    361 CTGCTGATGGCTCACTTTGATGTGGTGCCTGCCGCCCCTGAAGAA-GGCTGGGAGGTGCCCCC 419
          ||||||||||||||||||||||||||| ||||||||||||||||| ||||||||||||||||
Db    361 ATGCTGATGGCTCACATTGATGTGGTTCCTGCCCCGGAAGAACGGATGGGAGGTTGCCCCC 420

Qy    420 ATTCCTGTTGGGTTGGAGCGTGTAGCGAGCCCACACTGCGACACCACGAACGACAAGGAGAA 479
          ||||||| |||||||||||| | ||||| |||||||||||||||||||||||||||
Db    421 GTTCCTCAGCGCTGAACGACAATGGCTTCAACTTCAGGGCCGGGGTCCGCGTGGAGCAACARAAAA 480

Qy    480 CTCTGTGATGGCATTACTGCCAGGCGTCCTGCTGTCATCAGGAAGTACATCCCCCG 539
          |||| ||||||||||||||||||||||||||||||||||||||||||| || || |
Db    481 CTCTGTGAAGGGCGAATCCTGGAGCTCCTGTTGATCAGAAACTACA-GCCCAA 539

Qy    540 AAGAATCTTTCTTCATTTCTCTGGGCCCATGAT----GAGGAGTCATCAGGGACCAGGGGGCTC 595
          |||||||||||| ||| |||||||||||||| ||   |||||||||||||||||||||||||||
Db    540 AAGATCTTTCATTGCT-TGGGCCATGATGACGGAGTGTCCGGGGAAAAAGGGGGCTC 598

Qy    596 AGAGGATCTCAGCACTCTTACAGTCCACCTAGGGTCTCCAACTAGGCCTTCCTTGTTGGATGAC 658
          |||||||| | ||||||||||||||| ||| |||||||| |||||||||||||||||||||||||
Db    599 AGAAGATCTCAGCACTCTTACAGGGTGTCCAACTAGCCTTCCTTGTTGGATGAC 658

Qy    653 AGGGGGGGCTTCATCTTGGATGATTTCATTCCTAACTTCAAGAGAGCCCATCGCCTTGATTG 712
          |||| | ||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db    659 AGGGAGCTTCAACTTGGGAGGCGTTCATTCCAAACTCGAGAAACCAGTTGGCATGAATTT 718

Qy    713 CAGTCTCAGAGAGAAGGGTTCCATGAACC 739
          ||| | |||||||||||| | ||||||||
Db    719 CAGCCCACTGAGAACGTGGCCTTGAACC 745
```

```
Db             719 CAGCCCACTGAGAACGTGGCCTTGAACC 745
               ||| | |||||||||||| | ||||||||
               CAGTCTCAGAGAGAAGGGTTCCATGAACC 745
```

RESULT 14
LOCUS   BB613230
DEFINITION   BB613230 RIKEN full-length enriched, 10 day neonate skin Mus
             musculus cDNA clone 4732466D17 5', mRNA sequence.
ACCESSION   BB613230
VERSION   BB613230.1 GI:16453908
KEYWORDS   EST.
SOURCE   house mouse.
ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 666)
AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
             ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.
             , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
             ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.
             Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
             ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
             Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
             Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp;
             URL:http://genome.gsc.riken.go.jp/,
             Carninci,P., Shibata,K., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
             ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
             Watahiki,A., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
             ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
             Hayashizaki,Y.
             RIKEN Integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
             ,Y. and Hayashizaki,Y.
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
             ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
             Hayashizaki,Y.
             Computational Analysis of Full-Length Mouse cDNAs Compared with
             Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
             Please visit our web site (http://genome.gsc.riken.go.jp) for
             further details.

FEATURES   Location/Qualifiers
        source   1..666
                     /organism="Mus musculus"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="4732466D17"
                     /clone_lib="RIKEN full-length enriched, 10 day neonate
                     skin"
                     /sex="mixed"
                     /tissue_type="skin"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /note="Site_1: SalI; Site_2: BamHI; cDNA library was
                     e mouse tissues.
```

RESULT 15
LOCUS       BG608571              580 bp    mRNA    linear   EST 17-APR-2001
DEFINITION  307199 MARC 1PIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION   BG608571

VERSION     BG608571.1  GI:13658549
KEYWORDS    EST.
SOURCE      pig.
  ORGANISM  Sus scrofa
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE   1  (bases 1 to 580)
  AUTHORS   Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
            Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
            and Keele,J.W.
  TITLE     Design and use of two pooled tissue normalized cDNA libraries for
            EST discovery in swine
  JOURNAL   Unpublished (2000)
COMMENT     Contact: Smith TPL
            USDA, ARS, US Meat Animal Research Center
            PO Box 166, Clay Center, NE 68933-0166, USA
            Tel: 402 762 4366
            Fax: 402 762 4390
            Email: smith@email.marc.usda.gov
            Single pass sequencing. Bases called and alt_trimmed with phred
            v0.980904.e. Vector identified by cross_match with the -minscore 18
            and -minmatch 12 options.
            PCR Primers
            FORWARD: AGGAAACAGCTATGACCAT
            BACKWARD: GTTTTCCCAGTCACGACG
            Plate: 95  row: E  column: 10
            Seq primer: ATTTAGGTGACACTATAG.
FEATURES             Location/Qualifiers
     source          1..580
                     /organism="Sus scrofa"
                     /db_xref="taxon:9823"
                     /clone_lib="MARC 1PIG"
                     /tissue_type="pooled"
                     /lab_host="DH10B"
                     /note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
                     Library made from pooled tissue from day 11, 13, 15, 20,
                     and 30 embryos."

BASE COUNT       135 a     160 c     144 g     141 t
ORIGIN

Query Match
Best Local Similarity   27.2%;     Score 410.2;   DB 12;   Length 580;
Matches 488;   Conservative   0;   Mismatches  88;   Indels   5;   Gaps  2;

Qy   293 ATGAAGTCGTGGAAGAGTATAGCACCCACCTGTTCACTATCCAAGGCGTCGGAGACCCCACCTTGC 352
         ||||||||||||  ||| ||    |||| ||   ||| ||  || || ||  |  ||||  || |
Db     1 ATGAAGTCGTGGGGAGAG--TTAGCCCCGTTGACCTCCACTCGTCCACGGCGTCCGGACCCCAGCCTAC 58

Qy   353 AGCCCGTACCTGCTGATGGCGTCACTTTGATGTGGTGGCCTGCCCCCTGAAGAAGGCTGGGAGG 412
         |||||  || | | | | ||| |||  |||||||||| | ||| |||||||| |||||||| |
Db    59 AGCCCTACATCCTCGCTAGCCTCACCATTGATGTGGTGGTGGTTGCTGCCCCTGATGAAGGCTGGGACG 118

Qy   413 TGCCCCCATTCTCTGGGTTGGAGCGTGGATGGGCGTCATCTATGGTCGGGGCCACACTGGACG 472
         |||||| ||||||||||  |||| | | ||||| |||| |||||| |||| |||||| | ||||
Db   119 TGCCCCCTTTCTCTGGGCCTGGAACGGTAAATGGCTTTATCTACGGTCGGAGGCACAATAGACA 178

Qy   473 ACAAGAACTCGTGATGGGCATTACTGCCAGGCCTCCTGTGTGATCAGGAAGTACA 532
         ||||||||||  ||||||| | |||||||| | | |||   |||||||  | |||
Db   179 ACAAGAACTCTCTTATGGGAATCCTGCAGGCGCTTGGAACTTCTGCTCATCAGAAAACTACA 238

Qy   533 TCCCCCGAAGATACTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCAGGG---ACAG 589
         ||||  |||| |||  ||||||||| |||||||||||||||||||||  || |||   ||||
Db   239 TCCCTCGAAAGATCCTTTCTTCATTGCTCTGGGCCCATGATGAGGAGTTGATGGGGTCAACG 298

Qy   590 GGGCTCAGGAGATCTCAGCCCTGCTACAGTCAAGGGCGTGCCAGCTAGCCTTCATTGTGG 649
         |||||||| ||||||||||||| |||||| | ||||||||||||||||||||||||||||
Db   299 GGGCTCAGAAGATCTCAGCCCTTCTACAGCCGAAGGGCGTCCAGCTAGCCTTCATTGTGG 358

Qy   650 ACGAGGGGGCGGCTTCACCTTGGATGAGTTTCATTCTTCAAAGAAGCCCACCGCCTTGA 709
         | |||||| |   ||||| || |||| |||||||||  | ||| |||| |||  ||| |
Db   359 ATGAGGGGAGTTTCATCTTTGATGGTTTCATTCCCGCGCCTTAAGAAGAACCCCTTTGCCATGG 418

Qy   710 TTGCAGTCTCAGAGAAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAAACATGACTTCAGGCC 769

---

prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN. Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer [5'
GAGAGAGAAGGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. cDNA went through one round of normalization
to Rot = 10.0 and subtraction to Rot = 100.0. Second
strand cDNA was prepared with the primer adapter of
sequence [5' GAGAGAGATTCGGATTGAGTTAATTAAATTAATCCCCCCCCCC
3']. cDNA was cloned into the XhoI and BamHI sites.
Vector: a modified pBluescript KS(+) after bulk excision
from Lambda FLC I"

BASE COUNT       149 a     186 c     175 g     156 t
ORIGIN

Query Match
Best Local Similarity   27.9%;     Score 421.6;   DB 10;   Length 666;
Matches 512;   Conservative   0;   Mismatches 129;   Indels   3;   Gaps  1;

Qy     1 ATGGCTCAGCGGGTGCGTTTGCGTGCTGTGGCCGTGTGGGTGCTAATGGCTGCTAGTTTTGCCCT 60
         ||||| || |||| |   | ||
Db    21 ATGGCTGAGCTGACTTGCTACTTGCCCGCTGCTTGGGCAGCTGCTGGTGCTTCCTTCCTTTTTTCGCT 80

Qy    61 ACCGTCTCCAGATCGATGGGCCGGAGGAGCGGGGAGCATCAAAAGGGCGTCGCGAATCCCT 120
         || |||||| |||| | ||||| |||||| ||||| ||| |   ||||||| |||| |||||
Db    81 ACGGTCTCCGGATCCACTGGCCCTAGGAGCAGGGGGAAAATCGGGGGGCGTCCCGGATCCCT 140

Qy   121 TCTCAGTTCAGCAAAGAGGAAACGCGGCGTTCCGCGATGAGAAAGAGCGCGCTGAAAGGTTGCCATCCAG 180
         || |||||||||| |||||| | ||  ||| |||| | |  ||||||| |||||||||||| |||||||||
Db   141 TCCCAGTTCAGCGAGGAGGACCGTGTTCGCGTGCTATAAAAGAGGCGCGCTGAAAGGTGCCATCCAG 200

Qy   181 ATTCAAACACTGACTTTTTAGCTCTCGAGAAGTCCAAATACTACAGCGCCTCGGCGTAGGTTCGGA 240
         ||||  ||| |||| ||||||||| ||||||||| ||  |||||||||||| | |  |  || |||
Db   201 ATTCCCACAGTGTCTTTTCAGCCACGAGGAATCCAACACCACAGCCCCTTGCTGAGTTTTGGA 260

Qy   241 AAATACATTCATTAAAAGTCTTTCCTACAGTGGTCAGCACGCTTTATCCAGCCATGAAGTC 300
         |||| || | || |||| |||||||||||||| ||| |||| ||| |||| ||||||||||
Db   261 GAATATATCCGCAAAGCCTTCCCTACAGTGTTCACACGCAGCCTTGTCCAACATGAAGTC 320

Qy   301 GTGGAAGAGTATAGGCCACCTGTTCACTATATCCAAGGCGTGGAGTACATTCCCCGGA 360
         ||||  | ||||||||||||||||||   ||  ||||| || || |  |  || |||
Db   321 GTGGGCAAAGTATAGGCCACCTGTTCAAAAGGGCTGGAGGTGGTGGGGAGGCTGGAGG 380

Qy   361 CTGCTGATGGCGTCACTTTGATGTGGTGCCGTCCCCTTGAAAAAGGGTGTGGGAGGTGGCCCCCA 420
         |||||||||||||||| | ||||||||  | || ||| ||| |||| | |||| | |||||||
Db   381 ATGCTGATGGCGTCACAATTGATGTGGTTCCTCCCCCGGAAGAAGGATGGGAGGTGCCCCCG 440

Qy   421 TTCTCTGGGGTTGGGACGGTGATGGGCGGCGCATCTATGGTCGGACGACAAGAAC 480
         |||||  |  ||||| ||  | ||||| || |||| ||||||||  | | |||||
Db   441 TTCTCAGGCCGTGGAACGCAATGGGCTTCATCTTATGGCCGGGGGCTGCCCGGGGGGGCACACTGGACG 500

Qy   481 TCTGTGATGGATTACTGCCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATTCCCCGGA 540
         ||||||| ||  | | ||||| ||| ||||||||||  |||| |    | |||| || |||
Db   501 TCTGTGAAPTGGGGATCCGCATGCCTTTGGAGCTCCTGTTGATCAAAAAACTACAACCCCAA 560

Qy   541 AGATCTTGTCTTCATTTGTCTGGGCCATGATGAGGAGGAGTCATCAGGG----ACAGGGGCTCAG 597
         |||||| | ||||||| ||||||||| ||||||||||||||  | |||      ||||||||||
Db   561 AGATCTTTCTTCATTGCTTGTGGGCCCTGATGAGGAGGTGTCCCGGGAAAAGGGGCTCAA 620

Qy   598 AGGAGGTCAGCCCTGCTACAGTCAAGGGCGTTCCAGCTAGCCTT 641
         | | |  ||||| | ||||| | |||| || ||||  |  ||||
Db   621 AAAATCTCACACACTCTTACAGCGAGGGTGTGTTCCACTTACCTT 664

```
Db  419 TTGCAGTCTCCCAGAGAGGGCTTAATTAAACCTCATGCTGCAAGTAAACACGACTCCAGGCC 478
            ||||||||||| | ||||||||||||||||||||||||||||||| | |||| ||| |||||

Qy  770 ACTCTTCAGCTCCTCCAAAAGGAGGACAAGCAATTGGCATCCTTGCAGCTGCTGTCAGCCGAT 829
            |||||||||||||||||||||| ||||| | |||||| | |||||||| | |||||||| ||||
Db  479 ACTCTTCAGCTCCTCCAAAAGGAAACTAGCAATTGGTATCCTTACAGCCGCTGTCAACCGAC 538

Qy  830 TGGAGCAGAGCACCAATGCCCTATCATTATTTGGAAGCGGGACA 870
            |||| ||||||||||||||||  | ||| |||||||||  |||
Db  539 TGGAACAGAGACACCAATGCCAAACATGTTTGGAGAATGGGCCA 579
```

Search completed: July 1, 2003, 07:46:16
Job time : 2359 secs

OM nucleic - nucleic search, using sw model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: July 1, 2003, 05:37:45 ; Search time 85 Seconds
(without alignments)
5444.415 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 1509
Sequence: 1509
Sequence: 1 atggctcaagcggtgcgtttg..........ctcacctgcacaaactgtga 1509

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Scoring table: Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153383381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :   Issued Patents NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Match Length	DB	ID	Description
	1	49.2	3.3	1224	1	US-08-127-278-3	Sequence 3, Appl1
	2	49.2	3.3	1224	1	US-08-555-860-3	Sequence 3, Appl1
	3	46	3.0	1445	1	US-09-814-951A-1	Sequence 1, Appl1
	4	37.4	2.5	273	1	US-08-204-740-5	Sequence 5, Appl1
	5	37.4	2.5	273	3	US-09-081-167A-5	Sequence 5, Appl1
	6	37.4	2.5	273	3	US-09-081-395-5	Sequence 5, Appl1
	7	37.4	2.5	273	4	US-09-416-833-5	Sequence 5, Appl1
	8	37.4	2.4	273	5	PCT-US95-02521-5	Sequence 5, Appl1
	9	36.6	2.4	5517	1	US-09-120-663-1	Sequence 1, Appl1
c	10	34.8	2.3	3168	1	US-09-041-075A-24	Sequence 24, Appl1
c	11	34.8	2.3	3220	1	US-09-041-075A-19	Sequence 19, Appl1
c	12	34.8	2.3	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
c	13	34.8	2.3	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
	14	34.4	2.3	2065	4	US-09-129-668-5	Sequence 5, Appl1
	15	34.2	2.3	3542	4	US-09-149-476-63	Sequence 63, Appl1
	16	33.8	2.2	2690	1	US-08-524-757-11	Sequence 11, Appl1
	17	32.6	2.2	4403765	4	US-09-103-840A-2	Sequence 2, Appl1
	18	32.6	2.2	4411529	4	US-09-103-840A-1	Sequence 1, Appl1
	19	32	2.1	7218	1	US-08-232-463-14	Sequence 14, Appl1
	20	31.8	2.1	1221	1	US-09-347-878-10	Sequence 10, Appl1
	21	31.8	2.1	2436	1	US-09-318-448-27	Sequence 27, Appl1
	22	31.6	2.1	1861	3	US-09-129-668-7	Sequence 7, Appl1
	23	31.6	2.1	2265	4	US-08-906-865-2	Sequence 2, Appl1
	24	31.6	2.1	2436	3	US-09-129-668-2	Sequence 2, Appl1
	25	31.6	2.1	2413	4	US-09-613-182-10	Sequence 10, Appl1
	26	31.6	2.1	2265	2	US-08-968-542C-1	Sequence 1, Appl1
	27	31.4	2.1	84495	4	US-09-797-906-3	Sequence 3, Appl1
c	28	31.2	2.1	1083	4	US-09-116-498-7	Sequence 7, Appl1
c	29	31	2.1	420	1	US-08-470-179-112	Sequence 112, App
	30	31	2.1	2031	6	5495001-8	Patent No. 5495001
	31	31	2.1	3404	2	US-07-762-132A-1	Sequence 1, Appl1
	32	31	2.1	5816	4	US-09-220-641-4	Sequence 4, Appl1
	33	30.8	2.0	2839	5	PCT-US94-07297-38	Sequence 38, Appl
	34	30.8	2.0	3478	1	US-08-396-479B-1	Sequence 1, Appl1
	35	30.8	2.0	3478	1	US-08-818-823-1	Sequence 1, Appl1
	36	30.6	2.0	1863	3	US-09-199-229-1	Sequence 1, Appl1
	37	30.6	2.0	1863	4	US-09-443-087-1	Sequence 1, Appl1
	38	30.6	2.0	1863	4	US-09-687-298-1	Sequence 1, Appl1
c	39	30.6	2.0	4158	4	US-09-221-017B-249	Sequence 249, Appl
c	40	30.6	2.0	16836	4	US-09-147-236-1	Sequence 1, Appl1
c	41	30.6	2.0	16836	4	US-09-147-236-10	Sequence 10, Appl1
c	42	30.4	2.0	559	4	US-08-882-907-10	Sequence 10, Appl1
c	43	30.4	2.0	4403	2	US-08-284-941-1	Sequence 1, Appl1
c	44	30.4	2.0	4403	2	US-08-447-642-1	Sequence 1, Appl1
c	45	30.4	2.0	4403	4	US-09-236-503-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-127-278-3
Sequence 3, Application US/08127278
Patent No. 5498697
GENERAL INFORMATION:
APPLICANT: IWAKI, Kanso
APPLICANT: OHTA, Tsunetaka
APPLICANT: KURIOTO, Masahi
TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
TITLE OF INVENTION: PREPARATION OF SAID PROTEIN
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: BROWDY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/127,278
FILING DATE: 27-SEP-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 281136/1992
FILING DATE: 28-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: NEIMARK, Sheridan
REGISTRATION NUMBER: 20,520
REFERENCE/DOCKET NUMBER: IWAKI=2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1224 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1224

US-08-127-278-3

Query Match   3.3%;   Score 49.2;   DB 1;   Length 1224;

Best Local Similarity 51.9%; Conservative 0; Pred. No. 1.2e-05; Mismatches 123; Indels 3; Gaps 1;
Matches 136;

```
Qy    318 CCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCCTGATGGCTCACTT 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CGTGTTGACCTGGCCAGGCACCACCACCCTACACTCTCCTCCATCTTGCTCAACTCCACAC 242

Qy    378 TGATGTGGTGCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTT---GGA 434
          |||||||||||| ||| ||||||| || |||
Db    243 GGATGTGGTGCCCTGGTCCCAGGCACCAACATTGGAGTCACGAACCCCTTGAGGCCTTCAAGGA 302

Qy    435 GCGTGAATGGGCGGTCACTCTATTGTGCGGGGCACACTGGACGACAAGAACTCTGTGATGGCATT 494
          |||||||||||||||||||||||||||||||||||||||||
Db    303 TTCTGAGGGGCTACACATCTATTGCCAGGGGTGCCCAGGGGTGCGTCAGCCATCCAGTA 362

Qy    495 ACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAAGAATCTTTCTTCAT 554
          |||||||||||||||||||||||||||||||||||||||
Db    363 CCTGGAAGCTGTGAGGAGGCTGAAGGTGGAGGGGCCACCGGTTCCCAGAACCAATCCACAT 422

Qy    555 TTCTCTGGGCCATGATGAGGAG 576
          |||||||||||||||||||||
Db    423 GACCTTTGTGCCTGATGAGGAG 444
```

RESULT 2
US-08-555-860-3
; Sequence 3, Application US/08555860
; Patent No. 5585474
; GENERAL INFORMATION:
; APPLICANT: IWAKI, Kanso
; APPLICANT: OHTA, Tsunetaka
; APPLICANT: KURIOTO, Masahi
; TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
; TITLE OF INVENTION: PREPARATION OF SAID PROTEIN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: BROWDY AND NEIMARK
;   STREET: 419 Seventh Street, N.W., Suite 300
;   CITY: Washington
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20004
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/555,860
;   FILING DATE: 13-NOV-1995
;   CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/127,278
;   FILING DATE: 27-SEP-1993
;   APPLICATION NUMBER: JP 281136/1992
;   FILING DATE: 28-SEP-1992
; ATTORNEY/AGENT INFORMATION:
;   NAME: NEIMARK, Sheridan
;   REGISTRATION NUMBER: 20,520
;   REFERENCE/DOCKET NUMBER: IWAKI-2
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-628-5197
;   TELEFAX: 202-737-3528
;   TELEX: 248633
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1224 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
;   NAME/KEY: CDS

; LOCATION: 1..1224
US-08-555-860-3

Query Match   3.3%;   Score 49.2;   DB 1;   Length 1224;
Best Local Similarity 51.9%;   Pred. No. 1.2e-05;
Matches 136;   Conservative 0;   Mismatches 123;   Indels 3;   Gaps 1;

```
Qy    318 CCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCCTGATGGCTCACTT 377
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    183 CGTGTTGACCTGGCCAGGCACCACCACCCTACACTCTCCTCCATCTTGCTCAACTCCACAC 242

Qy    378 TGATGTGGTGCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTT---GGA 434
          |||||||||||| ||| ||||||| || |||
Db    243 GGATGTGGTGCCCTGGTCCCAGGCACCAACATTGGAGTCACGAACCCCTTGAGGCCTTCAAGGA 302

Qy    435 GCGTGAATGGGCGGTCACTCTATTGTGCGGGGCACACTGGACGACAAGAACTCTGTGATGGCATT 494
          |||||||||||||||||||||||||||||||||||||||||
Db    303 TTCTGAGGGGCTACACATCTATTGCCAGGGGTGCCCAGGGGTGCGTCAGCCATCCAGTA 362

Qy    495 ACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAAGAATCTTTCTTCAT 554
          |||||||||||||||||||||||||||||||||||||||
Db    363 CCTGGAAGCTGTGAGGAGGCTGAAGGTGGAGGGGCCACCGGTTCCCAGAACCAATCCACAT 422

Qy    555 TTCTCTGGGCCATGATGAGGAG 576
          |||||||||||||||||||||
Db    423 GACCTTTGTGCCTGATGAGGAG 444
```

RESULT 3
US-09-814-951A-1
; Sequence 1, Application US/09814951A
; Patent No. 6387661
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE, AND USES THEREOF
; FILE REFERENCE: CL001179
; CURRENT APPLICATION NUMBER: US/09/814,951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1445
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-814-951A-1

Query Match   3.0%;   Score 46;   DB 4;   Length 1445;
Best Local Similarity 51.6%;   Pred. No. 0.00015;
Matches 131;   Conservative 0;   Mismatches 120;   Indels 3;   Gaps 1;

```
Qy    326 CTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCCTGATGGCTCACTTGATGTGG 385
          || ||| |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db    202 CTGACTATGGCACCAAACCCTACACTCTCCTCCATCTTGCTCAACTCCCACACGGAGTGG 261

Qy    386 TGCCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTT---GGACGGTGATG 442
          ||||| | |||| ||||||| |||
Db    262 TGCCTGTCTTCAAGGAACATTGGAGTCACGACCCCTTGAGGCCTTCAAGGAGATTCTGAGG 321

Qy    443 GCGTCACTCTATTGTGCGGGGCACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAGG 502
          |||||||||||||||||||||||||||||||||
Db    322 GCTACATCTATGCCAGGGGTGCCCAGGACATGAAGTCGTCAGCATCCAGTACCTGGAAG 381

Qy    503 CCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAAGATCTTTCTTCATTTCTCTGG 562
          || || |||
Db    382 CTGTGAGGAGGCTGAAGGTGGAGGGGCCACCGGTTCCCAGAACCAATCCACACATGACCTTTG 441

Qy    563 GCCATGATGAGGAG 576
          ||||||||||||||
Db    442 TGCCTGATGAGGAG 455
```

RESULT 4

US-08-204-740-5
Sequence 5, Application US/08204740
Patent No. 5753432
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allgretti & Witcoff, Ltd.
; STREET: 10 S. Wacker Drive, Suite 3000
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,740
; FILING DATE: 04-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 575432nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-715-1000
; TELEFAX: 312-715-1234
; TELEX: 910-221-5317
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-204-740-5

Query Match   2.5%; Score 37.4; DB 1; Length 273;
Best Local Similarity   55.8%; Pred. No. 0.032;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

Qy 315 CCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGANGGCTCA 374
        ||| ||| | ||| ||||| | ||||||| | ||||| | |||| | |||||||||
Db 108 CCATCTGCTGGGGCAAGCTAGGCAGCGACCCCCAGAAGAAAACGTGTGCATTTACGGGCA 167

Qy 375 CTTTGATGTNGGTGGCCT--GCCCCTGAAGAAGGCGTGGGGAGGTGCCCCCANTTCTCTGGGGTTG 432
        | || ||| | |||| | ||||| |||| || ||||| ||| || || | || |||||
Db 168 CCTGGACGTGCAGCCTGCCCCCTTGGAGGACGCGGGTGGGGACGCGAGGCCCTTCACCTTGGTG 227

Qy 433 GAGGGTGANGGCGTCATCTATGGGGCACACTGGACGACGACAAG 477
        |||||| || || | ||||| |||||| || ||| | |||
Db 228 GACGCGGGAAGGCAAGCTCGTATGGGAGAGGCCTCCACGGACGAATAAG 272

RESULT 5
US-09-081-167A-5
Sequence 5, Application US/09081167A
Patent No. 6083745
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant

; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/081,167A
; FILING DATE: 18-MAY-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 608374nan, Kevin E
; REGISTRATION NUMBER: 35,303
; REFERENCE/DOCKET NUMBER: 93,354-KK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-913-0001
; TELEFAX: 312-913-0002
; TELEX:
INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-081-167A-5

Query Match   2.5%; Score 37.4; DB 3; Length 273;
Best Local Similarity   55.8%; Pred. No. 0.032;
Matches 92; Conservative 0; Mismatches 71; Indels 2; Gaps 1;

Qy 315 CCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGANGGCTCA 374
        ||| ||| | ||| ||||| | ||||||| | ||||| | |||| | |||||||||
Db 108 CCATCTGCTGGGGCAAGCTAGGCAGCGACCCCCAGAAGAAAACGTGTGCATTTACGGGCA 167

Qy 375 CTTTGATGTNGGTGGCCT--GCCCCTGAAGAAGGCGTGGGGAGGTGCCCCCANTTCTCTGGGGTTG 432
        | || ||| | |||| | ||||| |||| || ||||| ||| || || | || |||||
Db 168 CCTGGACGTGCAGCCTGCCCCCTTGGAGGACGCGGGTGGGGACGCGAGGCCCTTCACCTTGGTG 227

Qy 433 GAGGGTGANGGCGTCATCTATGGGGCACACTGGACGACGACAAG 477
        |||||| || || | ||||| |||||| || ||| | |||
Db 228 GAGGGGGAAGGCAAGCTGTATGGGAGAGGCCTCCACGGACGAATAAG 272

RESULT 6
US-09-081-395-5
Sequence 5, Application US/09081395
Patent No. 6083746
; GENERAL INFORMATION:
; APPLICANT: Gudkov, Andrei
; APPLICANT: Kazarov, Alexander
; APPLICANT: Mazo, Ilya
; APPLICANT: Roninson, Igor B
; TITLE OF INVENTION: Methods for Identifying Genetic
; TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
; TITLE OF INVENTION: Growth in Cancer Cells
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: McDonnell Boehnen Hulbert & Berghoff
; STREET: 300 S. Wacker Drive, 32nd Floor
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

```
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/081,395
;     FILING DATE: 18-MAY-1998
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: No. 608746nan, Kevin E
;     REGISTRATION NUMBER: 35,303
;     REFERENCE/DOCKET NUMBER: 93,354-KK
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 312-913-0001
;     TELEFAX: 312-913-0002
;     TELEX:
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 273 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA

US-09-081-395-5

Query Match     2.5%;  Score 37.4;  DB 3;  Length 273;
Best Local Similarity    55.8%;  Pred. No. 0.032;
Matches 92;  Conservative   0;  Mismatches 71;  Indels   2;  Gaps   1;

Qy  315 CCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCA 374
        |||  |||||  ||  |  ||||||||||| ||||||||| ||||| || |||||| ||||||
Db  108 CCATCTGCTCGGGCAAGCTAGGCAGCGACCCCCAGAGAAGAAAACCGTGTGCATTTACGGGCA 167

Qy  375 CTTTGATGTGGTGCCT--GCCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTG 432
        ||| |||  ||  ||||   ||||||||||  |||||  |||  ||||||||||| ||| ||||
Db  168 CCTGGACGTGCAGCCTGCGCGCCCTGGACGAGGACAGGCGAGCCCTTCACCTTGGTG 227

Qy  433 GAGCGTGATGGCGTCATCTATGGGTCGGGGCACACTGGACGACAAG 477
        ||||||||||||||||||||||  |||||||||||| |||||||| ||
Db  228 GAGCGCGGAAGGCAAGCTGTTATGGGAGGAGGCTTCCACGGACGATAAG 272
```

```
RESULT 7
US-09-416-833-5
; Sequence 5, Application US/0941683
; Patent No. 6197521
; GENERAL INFORMATION:
;   APPLICANT: Gudkov, Andrei
;   APPLICANT: Kazarov, Alexander
;   APPLICANT: Mazo, Ilya
;   APPLICANT: Roninson, Igor B
;   TITLE OF INVENTION: Methods for Identifying Genetic
;   TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
;   TITLE OF INVENTION: Growth in Cancer Cells
;   NUMBER OF SEQUENCES: 13
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Allgretti & Witcoff, Ltd.
;     STREET: 10 S. Wacker Drive, Suite 3000
;     CITY: Chicago
;     STATE: Illinois
;     COUNTRY: USA
;     ZIP: 60606
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/416,833
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/204,740
```

```
;   FILING DATE:
;   ATTORNEY/AGENT INFORMATION:
;     NAME: No. 6197521nan, Kevin E
;     REGISTRATION NUMBER: 35,303
;     REFERENCE/DOCKET NUMBER: 93,354-C
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 312-715-1000
;     TELEFAX: 312-715-1234
;     TELEX: 910-221-5317
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 273 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA

US-09-416-833-5

Query Match     2.5%;  Score 37.4;  DB 4;  Length 273;
Best Local Similarity    55.8%;  Pred. No. 0.032;
Matches 92;  Conservative   0;  Mismatches 71;  Indels   2;  Gaps   1;

Qy  315 CCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCA 374
        |||  |||||  ||  |  ||||||||||| ||||||||| ||||| || |||||| ||||||
Db  108 CCATCTGCTCGGGCAAGCTAGGCAGCGACCCCCAGAGAAGAAAACCGTGTGCATTTACGGGCA 167

Qy  375 CTTTGATGTGGTGCCT--GCCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTG 432
        ||| |||  ||  ||||   ||||||||||  |||||  |||  ||||||||||| ||| ||||
Db  168 CCTGGACGTGCAGCCTGCGCGCCCTGGACGAGGACAGGCGAGCCCTTCACCTTGGTG 227

Qy  433 GAGCGTGATGGCGTCATCTATGGGTCGGGGCACACTGGACGACAAG 477
        ||||||||||||||||||||||  |||||||||||| |||||||| ||
Db  228 GAGCGCGGAAGGCAAGCTGTTATGGGAGGAGGCTTCCACGGACGATAAG 272
```

```
RESULT 8
PCT-US95-02521-5
; Sequence 5, Application PC/7US9502521
; GENERAL INFORMATION:
;   APPLICANT:
;     TITLE OF INVENTION: Methods for Identifying Genetic
;     TITLE OF INVENTION: Suppressor Elements and Genes Associated with Malignant
;     TITLE OF INVENTION: Growth in Cancer Cells
;   NUMBER OF SEQUENCES: 13
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US95/02521
;     FILING DATE:
; INFORMATION FOR SEQ ID NO: 5:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 273 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: cDNA

PCT-US95-02521-5

Query Match     2.5%;  Score 37.4;  DB 5;  Length 273;
Best Local Similarity    55.8%;  Pred. No. 0.032;
Matches 92;  Conservative   0;  Mismatches 71;  Indels   2;  Gaps   1;

Qy  315 CCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCA 374
        |||  |||||  ||  |  ||||||||||| ||||||||| ||||| || |||||| ||||||
Db  108 CCATCTGCTCGGGCAAGCTAGGCAGCGACCCCCAGAGAAGAAAACCGTGTGCATTTACGGGCA 167

Qy  375 CTTTGATGTGGTGCCT--GCCCCTGAAGAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTG 432
        ||| |||  ||  ||||   ||||||||||  |||||  |||  ||||||||||| ||| ||||
Db  168 CCTGGACGTGCAGCCTGCGCGCCCTGGACGAGGACAGGCGAGCCCTTCACCTTGGTG 227
```

```
Qy    433 GAGCGTGATGGCGTCATCTATGTGGTCGGGGGGCCACACTGGACGACAAG 477
          |||| || ||| ||  ||  ||||||  ||  ||||  |||||| |||
Db    228 GAGCGGGAAGGCAAGCTGTGATGGGAGAGGGCTCCACGGACGATAAG 272

RESULT 9
US-09-120-663-1
; Sequence 1, Application US/0912O663
; Patent No. 6228644
; GENERAL INFORMATION:
;   APPLICANT: Bogdanove, Adam J.
;   APPLICANT: Kim, Jihyun Francis
;   APPLICANT: Wei, Zhong-Min
;   APPLICANT: Beer, Steven V.
;   TITLE OF INVENTION: HYPERSENSITIVE RESPONSE ELICITOR FROM
;   TITLE OF INVENTION: ERWINIA AMYLOVORA, ITS USE, AND ENCODING GENE
;   NUMBER OF SEQUENCES: 5
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
;     STREET: P.O. Box 1051, Clinton Square
;     CITY: Rochester
;     STATE: New York
;     COUNTRY: U.S.A.
;     ZIP: 14603
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Patentin Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/120,663
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US 60/055,105
;     FILING DATE: 06-AUG-1997
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Goldman, Michael L.
;     REGISTRATION NUMBER: 30,727
;     REFERENCE/DOCKET NUMBER: 19603/1661
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (716) 263-1304
;     TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 5517 base pairs
;     TYPE: nucleic acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: DNA (genomic)
US-09-120-663-1

Query Match             2.4%;  Score 36.6;  DB 4;  Length 5517;
Best Local Similarity   57.4%;  Pred. No. 0.41;
Matches  66;  Conservative  0;  Mismatches  49;  Indels  0;  Gaps  0;

Qy    235 TTCGGAAAATACATTCATAAAGTCTTTTCCTACAGTGGTCACCACCAGCTTTATCCAGCAT 294
          || ||  |||| |||  ||||| | ||| | ||||||| |||| |||| ||  |||||||
Db   1000 TTGGGTAAAGACACCCAGCACTATCTCGGCGACCACCATCCCAGCGACGGGTAGCCACCAT 1059

Qy    295 GGAGTCGTTGGAAGAGTATAGCCACACCTCTTCACTATTCCAAGGCTCGGACCCCAGCT 349.
          |  ||||  |||||  |||||||||||| |||||  ||  |||||  ||| | |||||
Db   1060 CTGCTGCTGGGACAACAAAAGGCACCCGTTGTTGATATCAAAAAGCACCGCCACCAGCT 1114

RESULT 10
US-09-041-075A-24
; Sequence 24, Application US/0904107SA
; Patent No. H002022
; GENERAL INFORMATION:
;   APPLICANT: Heidler, Steven A
;   APPLICANT: Radding, Jeffrey A
;   TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
```

```
; FILE REFERENCE: X-11242 Sequence List
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 24
;   LENGTH: 3168
;   TYPE: RNA
;   ORGANISM: Candida neoformans
US-09-041-075A-24

Query Match             2.3%;  Score 34.8;  DB 1;  Length 3168;
Best Local Similarity   35.4%;  Pred. No. 1.1;
Matches  73;  Conservative  26;  Mismatches 107;  Indels  0;  Gaps  0;

Qy   1074 CACAGTCAACTTCCGGATTCACCCTGGACAGAGACAGTCCAAGAGAGTCCTAGAACTCACGAA 1133
          ||| :|||  ||  |  ||| :||  ||| |  ||| |  | ||| |  :|  :   ||||
Db   1475 CCCACUCAAGCUCGGGAUACCCAUCGCAUACAUCGCCGUGUCAUCUCCGAUCACGUC 1534

Qy   1134 GAACAUUGUGCUGAUAACAGAGUUCCAGUUCCAUGUGUUGAUUGCCUUUGAUCCCUCC 1193
          || | :|:   ||| ||   |  | |:|  | ||: |||:  |||  |||||||||
Db   1535 GCAGUUGUCUGGCCGGGCGCACCCAUCUUUGCAUGGCUCAUCACCUUUUUCUCGCCCG 1594

Qy   1194 CGTCACCCCTTCGTGATGACAAGGCCGTTGGCCTACCAGCTGCTCCCCGAGACCGTACAGTC 1253
          | ||| ||||:  |||  ||  |||||  |  |  |||:  ||  |  | : | |||:
Db   1595 UUUCACCCCUUGCCGGGCGCGCCCCGAAAUCCACGUGCCUUCCCCCAGCCCUUGAAUC 1654

Qy   1254 CGTCTTCCCGGAAGTCAAATATTACTG 1279
          ||:|  |:|    | | |||  |:|:|
Db   1655 CGUCCUCUCUACGUGCCAAACACUUCUG 1680

RESULT 11
US-09-041-075A-19
; Sequence 19, Application US/0904107SA
; Patent No. /H002022
; GENERAL INFORMATION:
;   APPLICANT: Heidler, Steven A
;   APPLICANT: Radding, Jeffrey A
;   TITLE OF INVENTION: IPC SYNTHASE GENES FROM FUNGI
; FILE REFERENCE: X-11242 Sequence List
; Patent No. H002022
; CURRENT APPLICATION NUMBER: US/09/041,075A
; CURRENT FILING DATE: 1998-03-10
; PRIOR APPLICATION NUMBER: 60/043,591
; PRIOR FILING DATE: 1997-04-15
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 19
;   LENGTH: 3220
;   TYPE: DNA
;   ORGANISM: Candida neoformans
;   FEATURE:
;     NAME/KEY: intron
;     LOCATION: (1888)..(1939)
US-09-041-075A-19

Query Match             2.3%;  Score 34.8;  DB 1;  Length 3220;
Best Local Similarity   48.1%;  Pred. No. 1.1;
Matches  99;  Conservative  0;  Mismatches 107;  Indels  0;  Gaps  0;

Qy   1074 CACAGTCAACTTCCGGATTCACCCTGGACAGAGACAGTCCAAGAGAGTCCTAGAACTCACGAA 1133
          || ||||  ||  |  ||| |||  ||| | ||| | | |||||| | ||| |   ||||
Db   1475 CCCACTCAAGCTCGGGATAACCAGTCCATCGCATACATCGCGCGTGCATCTTCCCGATCACGTC 1534

Qy   1134 GAACATTGTGGGTGATAACAGAGTTCAGTTCCATGTGTTGACTGGCTTTGACCCCTCC 1193
          |||  |||| | |||  |  | | |||||| | | :  || ||| |||||| ||||
Db   1535 GCAGTTGTCTCGGCCGGGCGGACACCCATCTTTGCATGGCTCATCACCTTTTTTCTCGCCCG 1594

Qy   1194 CGTCAGCCCTTCTGATGACAAGGCCCTTGGGCTACCAGCTGCTCGCGAGACCGTACAGTC 1253
          | |||  ||  :  |||  | | | |  |  |  |||  ||  |  | | ||||||||
Db       (continued)
```

```
Db    1595 TTCATCCCTTCCGGGAAGTCAATATATTACTG 1279
           |||   ||||||||  ||||||||||||  ||||
Qy    1254 CGTGTTCCCGGAAGTCACGTGCCCGGAAATCCAGTGCCCTTGAATC 1654
           |||||||||| || |||||| |||||||||||||||| || |||
Db    1655 CGTCCTCTACGGTGCCAACATTTCTG 1680
```

Qy   1254 Sequence 2, Application US/0910384OA
Db   Patent No. 6294328
           GENERAL INFORMATION:
;   APPLICANT: FLEISCHMAN, Robert D.
;   APPLICANT: WHITE, Owen R.
;   APPLICANT: FRASER, Claire M.
;   APPLICANT: VENTER, John C.
;   SEQ ID NO 2
;   LENGTH: 4403765
;   TYPE: DNA
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: CDC 1551
;   OTHER INFORMATION: "n" bases at various positions throughout the sequence
;   US-09-103-840A-2

RESULT 12
US-09-103-840A-2/c
; Sequence 2, Application US/0910384OA
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence

Query Match   2.3%;   Score 34.8;   DB 4;   Length 4403765;
Best Local Similarity   54.8%;   Pred. No. 58;
Matches 69; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

```
Qy    361 CTGCTGATGGGCTCACTTTGATGTGGTGGTCCCTGCCCCCTGAAGAAGGCTGGGAGGTGCCCCCA 420
          |||| |  |||| ||||||   ||| ||||| ||||| |||||||||||| |||  ||| |||||
Db 2400010 CTGATCCACGGGCCACTCGATGTGGTGCCGGCTGAACCGGCGAATGGAGTGGTTGCACCCG 2399951
```

RESULT 13
US-09-103-840A-1/c
; Sequence 1, Application US/0910384OA
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; SOFTWARE: Patentin Ver. 2.1
; NUMBER OF SEQ ID NOS: 2
; CURRENT FILING DATE: 1998-06-24
; CURRENT APPLICATION NUMBER: US/09/103,840A
; FILE REFERENCE: 24366-20007.00
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS

```
Qy    421 TTCTCTGGGTTGGAGGCGTGATGAGGCGTCATCTATGGTGTCGGGGGCCACACTGGACGACAAGAAC 480
          |||| || |||  ||  |||||| |||||||||||||||||||||||||||| ||||| |||| ||
Db 2399950 TTTTCCGGCGCCATCGAAGACGGCTACGTCTGGGGTTCGTCGGCGGTCGGACGTGACATGAAGGAC 2399891
```

Qy   481 TCTGTG 486
          ||||||
Db 2399890 ATGGTG 2399885

RESULT 14
US-09-129-668-5
; Sequence 5, Application US/0912966B
; Patent No. 6429010
; GENERAL INFORMATION:
; APPLICANT: Greengard, Paul
; APPLICANT: Porton, Barbara
; APPLICANT: Kao, Hung-Teh
; SEQ ID NOS: 8
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin Ver. 2.0
; TYPE: DNA
; ORGANISM: Homo sapiens
; LENGTH: 2065
; US-09-129-668-5

Query Match   2.3%;   Score 34.4;   DB 4;   Length 2065;
Best Local Similarity   47.6%;   Pred. No. 1.1;
Matches 101; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

```
Qy   361 CTGCTGATGGGCTCACTTTGATGTGGTGGTCCCTGCCCCCTGAAGAAGGCTGGGAGGTGCCCCCA 420
         ||| ||||  ||||  || || |  || || | |||| || || |||||||| | |||| ||||||
Db 2401468 CTGATCCACGGGCCATCTCGATCCGATGTGGTGTCCGATGTGGTGCCGGCGCGCTGAACCGGCGAATGGAGTG 2401409
```

RESULT 15
US-09-149-476-63
; Sequence 63, Application US/0914947B
; Patent No. 6420526
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 186 Human Secreted proteins
; TITLE OF INVENTION:
; FILE REFERENCE: PZ002P1
; CURRENT APPLICATION NUMBER: US/09/149,476
; CURRENT FILING DATE: 1998-09-08
; EARLIER APPLICATION NUMBER: PCT/US98/04493

EARLIER FILING DATE: 1998-03-06
EARLIER APPLICATION NUMBER: 60/040,162
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,333
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/038,621
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,626
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,334
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,336
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/040,163
EARLIER FILING DATE: 1997-03-07
EARLIER APPLICATION NUMBER: 60/047,600
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,615
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,597
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,502
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,633
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,592
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,503
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,618
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,617
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,583
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,584
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,500
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,587
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,492
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,581
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,598
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,613
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,582
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,596
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,612
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,632
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,601
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/043,580
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,568
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,314
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,569
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,311
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,671
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,674

EARLIER APPLICATION NUMBER: 60/043,669
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,312
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,313
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,672
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/043,315
EARLIER FILING DATE: 1997-04-11
EARLIER APPLICATION NUMBER: 60/048,974
EARLIER FILING DATE: 1997-06-06
EARLIER APPLICATION NUMBER: 60/056,886
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,877
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,662
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,889
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,893
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,630
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,878
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,872
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,882
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,637
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,903
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,911
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,888
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,636
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,879
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,880
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,894
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,874
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,910
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,864
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,631
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,845
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/056,892
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/057,761
EARLIER FILING DATE: 1997-08-22
EARLIER APPLICATION NUMBER: 60/047,595
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,599
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,588
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,585
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,586
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,590
EARLIER FILING DATE: 1997-05-23
EARLIER APPLICATION NUMBER: 60/047,594

```
;       EARLIER FILING DATE: 1997-05-23
;       EARLIER APPLICATION NUMBER: 60/047,589
;       EARLIER FILING DATE: 1997-05-23
;       EARLIER APPLICATION NUMBER: 60/047,593
;       EARLIER FILING DATE: 1997-05-23
;       EARLIER APPLICATION NUMBER: 60/047,614
;       EARLIER FILING DATE: 1997-05-23
;       EARLIER APPLICATION NUMBER: 60/043,578
;       EARLIER FILING DATE: 1997-04-11
;       EARLIER APPLICATION NUMBER: 60/043,576
;       EARLIER FILING DATE: 1997-04-11
;       EARLIER APPLICATION NUMBER: 60/047,501
;       EARLIER FILING DATE: 1997-05-23
;       EARLIER APPLICATION NUMBER: 60/043,670
;       EARLIER FILING DATE: 1997-04-11
;       EARLIER APPLICATION NUMBER: 60/056,632
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,664
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,876
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,881
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,909
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,875
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,862
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,887
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/056,908
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/048,964
;       EARLIER FILING DATE: 1997-06-06
;       EARLIER APPLICATION NUMBER: 60/057,650
;       EARLIER FILING DATE: 1997-09-05
;       EARLIER APPLICATION NUMBER: 60/056,884
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/057,669
;       EARLIER FILING DATE: 1997-09-05
;       EARLIER APPLICATION NUMBER: 60/056,656
;       EARLIER FILING DATE: 1997-08-22
;       EARLIER APPLICATION NUMBER: 60/049,610
;       EARLIER FILING DATE: 1997-06-13
;       EARLIER APPLICATION NUMBER: 60/061,060
;       EARLIER FILING DATE: 1997-10-02

Query Match           2.38;    Score 34.2;  DB 4;  Length 3542;
Best Local Similarity   49.2%;  Pred. No. 1.9;
Matches   90;  Conservative   0;  Mismatches   93;  Indels   0;  Gaps   0;

Qy      1321 ACAAACCTCACCACTGGGCATCTACAGGTTCTACCCCATCTACATACAGCCTGAAGACTTC 1380
             | |||     | |||||||     ||  ||| ||| ||    | |||||      |    ||
Db      2477 ACACTGCACACCACCACGCGTCCAACTGGCTGCACCTCATCCCCAGACGCGCTGAGCACCTC 2536

Qy      1381 AAACGCATCCATGGAGTCAACGAGAGAAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAAA 1440
             || ||||||  |  ||||| ||| || ||||| |||  | | || | ||||| ||| |
Db      2537 AAGCGCACCGTGGAGAATATCAAGGATCCTTTGTTCAGGTTCTTTGAGAGAGAAGTGAAG 2596

Qy      1441 TTCATCTTTGAGTTGATTCAGAATGCTGACACAGACCAGCCAGTTTCTCACCTGCAC 1500
             |  ||  ||  || | |||||| |||| |  | |  | || || || ||| | |||
Db      2597 ATGGGCGCAAAGCTGCTTCAGGACGTTCGCCAGACCTTGCAGATGTCGTCCAGGTGTGC 2656

Qy      1501 AAA 1503
             ||
Db      2657 GAA 2659
```

```
OM nucleic - nucleic search, using sw model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             July 1, 2003, 07:07:34 ; Search time 269 Seconds
                    (without alignments)
                    8327.258 Million cell updates/sec

Title:              US-10-014-896-1
Perfect score:      1509
Sequence:           1 atggctcagcggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table:      IDENTITY_NUC
                    Gapop 10.0 , Gapext 1.0

Searched:           1055720 seqs, 74224136 residues

Total number of hits satisfying chosen parameters:          2111440

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
        Published Applications NA:*
                 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
                 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
                 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
                 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
                 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
                 7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
                 8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
                 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
                10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
                11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
                12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
                13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
                14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	74	4.9	280	10	US-09-783-590-10310	Sequence 10310, A
c	2	46	3.0	1445	12	US-10-109-860-1	Sequence 1, Appli
c	3	42	2.8	2710	9	US-10-037-270-931	Sequence 931, App
c	4	41.6	2.8	1988	9	US-10-073-885-54	Sequence 54, Appl
c	5	41.6	2.8	1997	10	US-09-925-301-555	Sequence 555, App
c	6	40.8	2.7	351	10	US-09-960-352-8484	Sequence 8484, Ap
c	7	40.4	2.7	2659	9	US-10-073-885-29	Sequence 29, Appl
c	8	39.8	2.6	14536	9	US-10-160-758-6	Sequence 6, Appli
c	9	39.8	2.6	14536	9	US-10-160-758-7	Sequence 7, Appli
c	10	38.8	2.6	699	9	US-10-184-644-138	Sequence 138, App
c	11	38.8	2.6	699	9	US-10-184-634-138	Sequence 138, App
c	12	38.8	2.6	699	10	US-09-799-946-5	Sequence 5, Appli
c	13	37.4	2.5	273	10	US-10-123-155-298	Sequence 298, Ap
c	14	36.8	2.4	466	9	US-10-108-605-280	Sequence 280, App
c	15	36.6	2.4	708	9	US-10-123-155-298 .	Sequence 298, App
c	16	36.6	2.4	5517	10	US-09-879-248-7	Sequence 7, Appli
c	17	35.6	2.4	546	9	US-09-918-995-27597	Sequence 27597, A
c	18	35.4	2.3	1049	9	US-10-123-155-358	Sequence 358, App
c	19	35.2	2.3	734	9	US-10-184-644-458	Sequence 458, App
	20	35.2	2.3	734	9	US-10-184-634-458	Sequence 458, App
	21	35.2	2.3	1731	9	US-09-738-626-148	Sequence 148, App
	22	35.2	2.3	3309400	9	US-09-738-626-1	Sequence 1, Appli
	23	35	2.3	2560	9	US-10-045-202-1	Sequence 1, Appli
	24	34.8	2.3	3168	9	US-09-742-581-24	Sequence 24, Appl
	25	34.8	2.3	3168	9	US-09-742-582-24	Sequence 24, Appl
	26	34.8	2.3	3168	10	US-09-742-580-19	Sequence 19, Appl
	27	34.8	2.3	3220	9	US-09-742-581-19	Sequence 19, Appl
	28	34.8	2.3	3220	9	US-09-742-582-19	Sequence 19, Appl
	29	34.8	2.3	3220	10	US-10-122-805-5	Sequence 5, Appl
	30	34.4	2.3	2065	9	US-10-124-800-1	Sequence 9, Appli
	31	34.2	2.3	1278	9	US-10-184-644-312	Sequence 312, App
	32	34.2	2.3	3542	9	US-09-809-391-63	Sequence 63, Appl
	33	34.2	2.3	8730	9	US-10-124-800-1	Sequence 19, Appl
c	34	34	2.3	802	9	US-10-184-634-312	Sequence 312, App
c	35	34	2.3	802	9	US-10-153-668-413	Sequence 413, App
c	36	33.8	2.2	1344	9	US-10-153-668-301	Sequence 301, App
	37	33.8	2.2	2673	9	US-09-976-059-1	Sequence 1, Appli
	38	33.6	2.2	88421	10	US-09-771-208-20	Sequence 20, Appl
	39	33.6	2.2	1027	10	US-09-803-589-13	Sequence 13, Appl
	40	33.2	2.2	659158	10	US-09-815-242-9401	Sequence 9401, Ap
	41	33	2.2	1872	12	US-09-771-208-20	Sequence 20, Appl
	42	33	2.2	2026	10	US-10-044-090-648	Sequence 648, App
c	43	32.8	2.2	3316	9	US-10-242-056-58	Sequence 58, Appl
c	44	32.6	2.2	594	9	US-10-123-155-10	Sequence 10, Appl
c	45	32.4	2.1	430	10	US-09-960-352-15018	Sequence 15018, A
```

ALIGNMENTS

RESULT 1
US-09-783-590-10310/c
Sequence 10310, Application US/09783590
Patent No. US2002011085oA1
GENERAL INFORMATION:
;   APPLICANT: Dillon, Patrick J.
;   APPLICANT: Haseltine, William A.
;   APPLICANT: Li, Haodong
;   APPLICANT: Rosen, Craig A.
;   APPLICANT: Ruben, Steven M.
;   TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
;   FILE REFERENCE: PO-16.2C1
;   CURRENT APPLICATION NUMBER: US/09/783,590
;   CURRENT FILING DATE: 2000-02-15
;   PRIOR APPLICATION NUMBER: 08/420,856
;   PRIOR FILING DATE: 1995-04-12
;   PRIOR APPLICATION NUMBER: 08/346,731
;   PRIOR FILING DATE: 1994-11-21
;   NUMBER OF SEQ ID NOS: 12485
;   SOFTWARE: PatentIn Ver. 2.0
;
SEQ ID NO 10310
;   LENGTH: 280
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc feature
;   LOCATION: (7)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (14)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (44)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (85)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (111)
;   OTHER INFORMATION: n equals a,t,g, or c
;   NAME/KEY: misc feature
;   LOCATION: (151)

```
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (182)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (203)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (228)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (244)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (246)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (259)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (278)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-10310

  Query Match       4.9%;   Score 74;  DB 10;  Length 280;
  Best Local Similarity 96.1%;  Pred. No. 1.5e-14;
  Matches 74; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy   489 GGCATTACTGCAGGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAAGATCTTT 548
         |||||||||||||||||||||||||||||||||| |||||||||||||||||| |||||||
Db    77 GGCATTACTGCAGGGCCTTGGAGCTCCTGCTGATNAGGAAGTACATCCCCGAAAGATCTTT 18

Qy   549 CTTCATTTCTCTGGGGCC 565
         |||| ||||| | ||||||
Db    17 CTTNATTTCTNTNGGGCC 1

RESULT 2
US-10-109-860-1
; Sequence 1, Application US/10109860
; Patent No. US20020142421A1
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al.
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE, AND USES THEREOF
; FILE REFERENCE: CL001179DIV
; CURRENT APPLICATION NUMBER:
; CURRENT FILING DATE: 2002-04-01
; PRIOR APPLICATION NUMBER: US/10/109,860
; PRIOR FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 1445
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-109-860-1

  Query Match       3.0%;   Score 46;   DB 12;  Length 1445;
  Best Local Similarity 51.6%;  Pred. No. 0.0017;
  Matches 131; Conservative 0; Mismatches 120; Indels 3; Gaps 1;

Qy   326 CTATCCAAGGCTCGGACCCCAGCGTTGCAGCCCTACCTGCTCACTTTGATGTGG 385
         ||| ||  | ||| || ||||  |  | ||||| || |  |  |   |   |||
Db   202 CTGACTATGGCACCAACACCCTACACTCTCTGCTCAACTCCACACGGATGTGG 261

Qy   386 TGCCTGCCCCCTGAACAAGAGGCTGGGAGGTGCCCCCATTCTCTGGGTT---GGAGGTGATG 442
         ||||||  |  |  | ||  |   |  |||||   || |||||| |||||     | || |||
Db   262 TGCCTGTCTTCAAGGAAACATTGGAGTCACGAACCCCCTTTGAGGCTTCAAGGATTCTGAGG 321

Qy   443 GCGTCATCTATGGTCGGGGGGCCACACTGACGACGAACAGAACTTCTGTGTAGGCGG 502
         || ||ATCTATG  | ||||| ||| ||| ||||  |  | |  || |   ||| ||
Db   322 GCTACATCTATGCCAGGGGGTGCCCAGGACATGAAGTGCGTCAGCGATCCAGTGATGCCAGTACCTGGAAG 381
```

```
Qy   503 CCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAAGATCTTTCTTCATTTCTCTGG 562
         | ||| ||  | || || ||  || |  |  ||| ||  |  | |   |   |   ||
Db   382 CTGTGAGGAGGCTGAAGGTGGAGGGCCCACCGGTTCCCCAGAACCATCCACATGACCTTTG 441

Qy   563 GCCATGATGAGGAG 576
         | | ||||||||||
Db   442 TGCCTGATGAGGAG 455

RESULT 3
US-10-037-270-931
; Sequence 931, Application US/10037270
; Publication No. US20030104529A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; APPLICANT: Liu, Chenghua
; APPLICANT: Asundi, Vinod
; APPLICANT: Zhang, Jie
; APPLICANT: Ren, Feiyan
; APPLICANT: Chen, Rui-hong
; APPLICANT: Zhao, Qing A.
; APPLICANT: Wehrman, Tom
; APPLICANT: Xue, Aidong J.
; APPLICANT: Yang, Yonghong
; APPLICANT: Wang, Jian-Rui
; APPLICANT: Zhou, Ping
; APPLICANT: Ma, Yunqing
; APPLICANT: Wang, Dunrui
; APPLICANT: Wang, Zhiwei
; APPLICANT: Tillinghast, John
; APPLICANT: Drmanac, Radoje T.
; TITLE OF INVENTION: No. US20030104529A1el Nucleic Acids and
; TITLE OF INVENTION: Polypeptides
; FILE REFERENCE: 784CIP2B
; CURRENT APPLICATION NUMBER: US/10/037,270
; CURRENT FILING DATE: 2002-01-04
; PRIOR APPLICATION NUMBER: 09/552,317
; PRIOR FILING DATE: 2000-04-25
; PRIOR APPLICATION NUMBER: 09/488,725
; PRIOR FILING DATE: 2000-01-21
; NUMBER OF SEQ ID NOS: 1104
; SOFTWARE: pt_FL_genes Version 1.0
; SEQ ID NO 931
;   LENGTH: 2710
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (218)..(1645)
US-10-037-270-931

  Query Match       2.8%;   Score 42;   DB 9;   Length 2710;
  Best Local Similarity 54.6%;  Pred. No. 0.0062;
  Matches 106; Conservative 0; Mismatches 85; Indels 3; Gaps 1;

Qy   319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTCACTTT 378
         ||| ||  | || ||| |  | ||||| |  ||||| |  |  |  |  |||||
Db   458 CTGCTCGGCAGGCTGGGCTCTGACCCACAGAAGAAGACCGTGTGCATTTACGGACCACCTG 517

Qy   379 GATGTGGGTGCCT---GCCCCTGAAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTGGAG 435
         |||||| | ||||    ||||||||||||  | |  | ||    ||  |  | | ||| |||
Db   518 GATGTGCAGCCCTGCAGCCCTGGGAGGACGGGCCACACTGGACGACGAAGAAACTCGTGATGGGGCATTA 577

Qy   436 CGTGATGGGGTCATCTATGGTCGGGGGGCACACTGACGACGAAGAAACTCGTGATGGGGCATTA 495
         || || ||| ||  | |||   | ||||| | |  |    |||| |  ||| | |||||| ||||
Db   578 CGAGACGGCGAAGCTTGTATGGGAGGAGGTTGGACTGATGATNAAGGGCCGGGTGGCCGGGTGG 637

Qy   496 CTGGCAGGCCTTGGA 509
         ||||||||| |||||
Db   638 ATAAACGCCCTGGA 651
```

RESULT 4
US-10-073-885-54
Sequence 54, Application US/10073885
Publication No. US20030096246A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PJZ03C1
; CURRENT APPLICATION NUMBER: US/10/073,885
; CURRENT FILING DATE: 2002-02-14
; Prior Application removed - See file Wrapper or Palm
; NUMBER OF SEQ ID NOS: 116
; SOFTWARE: Patentin Ver. 2.0
US-10-073-885-54
; SEQ ID NO 54
; LENGTH: 1988
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1898)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1901)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1908)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1980)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1985)
; OTHER INFORMATION: n equals a,t,g, or c

Query Match   2.8%;   Score 41.6;   DB 9;   Length 1988;
Best Local Similarity   54.1%;   Pred. No. 0.0069;
Matches 105; Conservative 1; Mismatches 85; Indels   3; Gaps   1;

Qy   319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTTGATGGCTCACTTT 378
              ||| || ||| |||||| ||||| || |||||||||| | ||||
Db   394 CTGCTCGGCCAGGCTGGGCTCCGACCCACAGAAGAAGAGACCCGTTGCATTTACGGGCACCTG 453

Qy   379 GATGTGTGTGCCT---GCCCCTGAAGAAGGCTGGGGAGGTGCCCCCAATTCTCTGGGTTGGAG 435
              |||||| ||| |||||||| ||||||||||||||||||||||||||| ||||||||
Db   454 GATGTGCAGCCTGCACGCCCTGGAGGACGGCGTGGGACAGGCCCCTTCACCCTGTGGAG 513

Qy   436 CGTGATGGCGTCATCTATGGTGTCGGGGCCACACTGCGACGACAAGAAACTCTGTGTCATTA 495
              | ||| | || || ||||| |||||||||||||||||||||||| ||| | ||||||
Db   514 CGAGACGGCAAGCTGTATGGGAGAGGTTCGACTGATGATAAGGGCCCGGGTGCCCGGCTGG 573

Qy   496 CTGCAGGCCTTGGA 509
              | ||| |||| ||||
Db   574 ATAAACGCCCTGGA 587

RESULT 5
US-09-925-301-555
Sequence 555, Application US/09925301
Patent No. US20020052308A1
GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA106
; CURRENT APPLICATION NUMBER: US/09/925,301
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05882
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1694
; SOFTWARE: Patentin Ver. 2.0
US-09-925-301-555
; SEQ ID NO 555
; LENGTH: 1997
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1887)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1951)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1980)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc_feature
; LOCATION: (1992)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-301-555

Query Match   2.8%;   Score 41.6;   DB 10;   Length 1997;
Best Local Similarity   54.1%;   Pred. No. 0.0069;
Matches 105; Conservative 1; Mismatches 85; Indels   3; Gaps   1;

Qy   319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTTGATGGCTCACTTT 378
              ||| || ||| |||||| ||||| || |||||||||| | ||||
Db   394 CTGCTCGGCCAGGCTGGGCTCCGACCCACAGAAGAAGAGACCCGTTGCATTTACGGGCACCTG 453

Qy   379 GATGTGTGTGCCT---GCCCCTGAAGAAGGCTGGGGAGGTGCCCCCAATTCTCTGGGTTGGAG 435
              |||||| ||| |||||||| ||||||||||||||||||||||||||| ||||||||
Db   454 GATGTGCAGCCTGCACGCCCTGGAGGACGGCGTGGGACAGGCCCCTTCACCCTGTGGAG 513

Qy   436 CGTGATGGCGTCATCTATGGTGTCGGGGCCACACTGCGACGACAAGAAACTCTGTGTCATTA 495
              | ||| | || || ||||| |||||||||||||||||||||||| ||| | ||||||
Db   514 CGAGACGGCAAGCTGTATGGGAGAGGTTCGACTGATGATAAGGGCCCGGGTGCCCGGCTGG 573

Qy   496 CTGCAGGCCTTGGA 509
              | ||| |||| ||||
Db   574 ATAAACGCCCTGGA 587

RESULT 6
US-09-960-352-8484
Sequence 8484, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
US-09-960-352-8484
; SEQ ID NO 8484
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 36-LIB34-055-Q1-E1-A8
US-09-960-352-8484

Query Match   2.7%;   Score 40.8;   DB 10;   Length 351;
Best Local Similarity   50.0%;   Pred. No. 0.0041;
Matches 130; Conservative 0; Mismatches 127; Indels   3; Gaps   1;

Qy   320 TGTTCACTATCCAAAGGGTCGGACCCAGCTTGCAGCCCTACCTGCTTGATGGCTCACTTTG 379
              |||| || | ||| | || ||||||||||| ||||| | |||||||||| |||
Db   83 TGTTAACCTGGCCCAGGCACCAACCCAAACTCCCTCCGTCTTGCTCAACTCCCATACGG 142

Qy   380 ATGTGGTTGCCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCCAATTCTCTGGGTT---GGAGC 436
              |||||||||||| |||| || | || | ||| |||| ||| || |||||
Db   143 ATGTGGGTTCCTGTCTTCCAGGAATATTGGAGTTCATGAGTCCCTTTTGAGGCCTTCAAGGATG 202

Qy   437 GTGATGGGCTCATCTATGGTCGGGGCCACACTGCGACGACAAGAAACTCTGTGTGGTCAGCA 496
              | || | | | ||||| |||||||||||||||||||||||| ||| || ||
Db   203 CAGACGGCTACACATCTATGGCAGGGGCGCCCAGGACATGAAGTGTGTCAGCATCCAAATACC 262

RESULT 7
US-10-073-885-29

Sequence 29, Application US/10073885
Publication No. US20030096346A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PJZ03C1
CURRENT APPLICATION NUMBER: US/10,073,885
CURRENT FILING DATE: 2002-02-14
PRIOR APPLICATION NUMBER: 116
Prior Application removed - See file Wrapper or Palm
NUMBER OF SEQ ID NOS: 116
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 29
LENGTH: 2659
TYPE: DNA
ORGANISM: Homo sapiens
US-10-073-885-29

Query Match   2.7%; Score 40.4; DB 9; Length 2659;
Best Local Similarity 54.1%; Pred. No. 0.021;
Matches 105; Conservative 0; Mismatches 86; Indels 3; Gaps 1;

Qy 319 CTGTTCACTATCCCAAGGCTCCGGACCCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTT 378
         ||| | ||| |||||||||| | |||| | ||||| ||||||| ||||||||||
Db 395 CTGCTCGGCAGCGTGCCCTCCGACCCACAGAGAAGGAAGAACCGTGGTCGATTCTGCATTTACGGGCACCTG 454

Qy 379 GATGTGGTGCCCTG---CCCCTGAAGAGAAGGCTCGGGAGGAGTGCCCCCAATTCTCTGGGTTGGGAG 435
         ||||| || || ||||||| |||||||| ||||| |||||||||||| ||| |||||||||
Db 455 GATGTGCAGCCTGCCAGCCCTGAGGACGGGCTGGGACACGGGGTGGGACGAGGACGGGCTGGGAG 514

Qy 436 CGTGGATGGCGTGCATCTATGGTGTCGGGGGGCCACACTGGACGACGACAAGAAACTCTGTGTCATGGCCATTA 495
         || |||||| |||| | ||||| |||||||| ||||| ||||| |||||| |||||||||||||
Db 515 CGAGACGGCAAGCTGTATGGGAGGAGTTCGACTGATGATAAGGCCCGGTGGCCGCGGCGGCTGG 574

Qy 496 CTGCAGGCCTTGA 509
         | | ||||| ||||
Db 575 ATAAAGCCCTGA 588

RESULT 8
US-10-160-758-6/c

Sequence 6, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10,160,758
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
CURRENT APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin version 3.1
SEQ ID NO 6
LENGTH: 14536
TYPE: DNA
ORGANISM: Homo sapiens
US-10-160-758-6

Qy 497 TGCAGGCCTTGGAGCTCCTGCTGATCAGGAAAGTACATCCCCCGAAAGAATCTTTCTTCATTT 556
         || ||| ||| ||| || | ||||||||| ||||||||||||||||||||||||
Db 263 TGGAGGCTGTGAGGAGGCTGAAGGCTGAAGGCCACCATTCCCAGACCATCCATTTGA 322

Qy 557 CTCTGGGCCATGATGAGGAG 576
         |||||||||||||||||||
Db 323 CCTTTGTGCCAGATGAGGAG 342

RESULT 9
US-10-160-758-7/c

Sequence 7, Application US/10160758
Publication No. US20030036076A1
GENERAL INFORMATION:
APPLICANT: EXELIXIS, INC.
TITLE OF INVENTION: CADs AS MODIFIERS OF THE p53 PATHWAY AND METHODS OF USE
FILE REFERENCE: EX02-089C
CURRENT APPLICATION NUMBER: US/10/160,758
CURRENT FILING DATE: 2002-06-03
PRIOR APPLICATION NUMBER: US 60/296,076
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: US 60/328,605
PRIOR FILING DATE: 2001-10-10
PRIOR APPLICATION NUMBER: US 60/357,253
PRIOR FILING DATE: 2002-02-15
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patentin version 3.1
SEQ ID NO 7
LENGTH: 14536
TYPE: DNA
ORGANISM: Homo sapiens
US-10-160-758-7

Query Match   2.6%; Score 39.8; DB 9; Length 14536;
Best Local Similarity 47.1%; Pred. No. 0.11;
Matches 122; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy   990 CAATGCAATAATCAGGACCACCACCACGGCCACTCACCACATATTCAAAAGCAGGGGTCAAGTTCAA 1049
          || || || ||||||| |||||||| || || |||| |||||||||||| ||||||||||
Db 8771 CACTGTTCTCAACCACACAGATCCTCTTGTACTTCTTCAGAAGCAAATCGGGAGGCCGGGTCGTGTCCGT 8712

Qy 1050 TGTCATCCCCCAGTGGCCCAGGCCACCAGTCAACTTCCGGATTCACCCTGGACAGACAGT 1109
          ||| |||| |||||||||||||| ||||||||||||| ||||| |||||||| || ||
Db 8711 TCTCATCTGTAATGAAGACCTGAACCAGGCGTCTCAGTTCAAGTTCCTGGA 8652

Qy 1110 CCAAGAGGTCCTTAGAACTCACGAAGAGAACAGTTGTTGGGCTGATAACAGAGTCCAGTTCCATGT 1169
          || || |||||||||| |||||||||| |||| ||| |||| ||||| ||||| || |||
Db 8651 GGTCATAGGCCACCACATGAAAATGATAAGGTCTGGCAGGTCTCACAGTCAAGTCATGGA 8592

Qy 1170 GTTGAGTGCCTTTGACCCCCTCCCCGTCAGCCCCTTCTGATGACAAGGCCTTGGGCTACCA 1229
          || || | ||| || ||| |||||||||||| ||||||| |||||||||||||||
Db 8591 GTGTGGTGATCCAACCACTCTCACTGTTCAATGGCAAAGAGCTCAATGGACAATTGCTACCAG 8532

Qy 1230 GCTGCTCCGCCAGACCGTA 1248
          || | |||||||||||||
Db 8531 GGTCTGCAGACAGCCTGTA 8513

RESULT 10
US-10-184-644-138/c
Sequence 138, Application US/10184644
; Publication No. US20030044930A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 138
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-644-138

Query Match　　　2.6%;　　Score 38.8;　　DB 9;　　Length 699;
Best Local Similarity　8.6%;　Pred. No. 0.031;
Matches　44;　Conservative 146;　Mismatches 323;　Indels　0;　Gaps　0;

Qy　945 ATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAG 1004
　　　　　|:|　　　|::|　|:::|　:　:|　||:|　|:　||||　|：||：|　||:|:|
Db　697 AW....DT.AN.BRTH.S..N.HHS......RM.A.C...S.KMHY...M..H.BNN.SSR 638

Qy 1005 GACCACCACGGGCACTCACCATATTCAAAGCAGGGGGTCAAGTTCAATGTCATCCCCCAGT 1064
　　　　　:|　|::|　:　|:::|　|:　||:|　|||:|　|::|　||||　||:|　|:|||:|
Db　637 CCS.BYA.YS.S....WCWH.RMYHCABYR.R..S.BA...Y...NDYHT.MB.BHYYC... 578

Qy 1065 GGCCCAGGCCACAGTCAACTTCCGGATTCACCCTGGACGAGACAGTCCAAGAGGTCCTAGA 1124
　　　　　||:|　||:|　|:::|　|：||:|　|:||:|　||:|　|::||:|:|　||::|
Db　577 R.N.DB....S.CA.K.YNK.RGA.K...T...KDST..CDRBC.T.AC...GB.DHKMTS. 518

Qy 1125 ACTCACGAAGAAACATGTGGCTGGATAAACAGAGTCCAGTTCCATGTGTTGGCCTTTGA 1184
　　　　　:|　||:|　:|::|　:|:||　|:||:|　||:|:|:|　|:||::|：||
Db　517 YCDT...CB.A.MS...SDY...DAYBC...T.BRH....A...MCC..W.M..R.RRSHTYTC. 458

Qy 1185 CCCCCTCCCCGTCAGCCCCTTCTGATGACAAGGCCTTGGGTACCAGCTGCTCCGCCAGAC 1244
　　　　　|||:|　|::|　|:||:|　|:||:|　|:||:|　||::|　|:||::|　|||:
Db　457 A.BC.DA...CSDAN..K.B.CHMMCS.CD.D.ABM.BT...MMHCA.YS..Y......SAB 398

Qy 1245 CGTACAGTCCGTCGTCTTCCCGGAAGTCAATATTACTGCCCCAGTTACTTCTATTGGCAACAC 1304
　　　　　:|:|　|||:|　:|:|：|　:|：||::|：|||：|　:||::|　|||
Db　397 .TBASHSAWY..DTBCAC.NBHSR..MS.AY.T.K.YNC.H.TMY...SHCS.DRTS..C 338

Qy 1305 AGACAGGCCGATTCTTTACAAAACCTCACCACTGGCCATCTACAGGTTCTCCCCATCTACAT 1364
　　　　　::|　|:|　|:||:|　|:||:|　|:||　|::|　|::|　|::|　|:||:|
Db　337 C..Y.AS...NBHBHM...D.MMSBWMYT....MY.C..WKAH.TH..A.C.C.W.S.SYS 278

Qy 1365 ACAGCCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAAAAATCTCAGTCCAAGCCTA 1424
　　　　　:|:|　|::|　|:||:|　|:||:|　||:|　|:||:|　|:||:|　|:||
Db　277 B.M..HDCWT.MYRCMWTD..BH.BCM.YRN.DB.AC.AYT......MA.B.TYYS...AC 218

Qy 1425 TGAGACCCAAGTGAAATTCATCTTTGAGTTGAT 1457
　　　　　:|:|　|:|　|:|　|||　||
Db　217 .....BYCYWS.BAY..HC..Y.HB.T..YMAT 185

RESULT 11

US-10-184-634-138/c
Sequence 138, Application US/10184634
; Publication No. US20030068884A1
; GENERAL INFORMATION:
; APPLICANT: Baker,Kevin P.
; APPLICANT: Chen,Jian
; APPLICANT: Desnoyers,Luc
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Pan,James
; APPLICANT: Smith,Victoria
; APPLICANT: Watanabe,Colin K.
; APPLICANT: Wood,William I.
; APPLICANT: Zhang,Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3430R1C217
; CURRENT APPLICATION NUMBER: US/10/184,634
; CURRENT FILING DATE: 2002-06-28
; PRIOR APPLICATION NUMBER removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 138
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-184-634-138

Query Match　　　2.6%;　　Score 38.8;　　DB 9;　　Length 699;
Best Local Similarity　8.6%;　Pred. No. 0.031;
Matches　44;　Conservative 146;　Mismatches 323;　Indels　0;　Gaps　0;

Qy　945 ATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAG 1004
　　　　　|:|　　　|::|　|:::|　:　:|　||:|　|:　||||　|：||：|　||:|:|
Db　697 AW....DT.AN.BRTH.S..N.HHS......RM.A.C...S.KMHY...M..H.BNN.SSR 638

Qy 1005 GACCACCACGGGCACTCACCATATTCAAAGCAGGGGGTCAAGTTCAATGTCATCCCCCAGT 1064
　　　　　:|　|::|　:　|:::|　|:　||:|　|||:|　|::|　||||　||:|　|:|||:|
Db　637 CCS.BYA.YS.S....WCWH.RMYHCABYR.R..S.BA...Y...NDYHT.MB.BHYYC... 578

Qy 1065 GGCCCAGGCCACAGTCAACTTCCGGATTCACCCTGGACGAGACAGTCCAAGAGGTCCTAGA 1124
　　　　　||:|　||:|　|:::|　|：||:|　|:||:|　||:|　|::||:|:|　||::|
Db　577 R.N.DB....S.CA.K.YNK.RGA.K...T...KDST..CDRBC.T.AC...GB.DHKMTS. 518

Qy 1125 ACTCACGAAGAAACATGTGGCTGGATAAACAGAGTCCAGTTCCATGTGTTGGCCTTTGA 1184
　　　　　:|　||:|　:|::|　:|:||　|:||:|　||:|:|:|　|:||::|：||
Db　517 YCDT...CB.A.MS...SDY...DAYBC...T.BRH....A...MCC..W.M..R.RRSHTYTC. 458

Qy 1185 CCCCCTCCCCGTCAGCCCCTTCTGATGACAAGGCCTTGGGTACCAGCTGCTCCGCCAGAC 1244
　　　　　|||:|　|::|　|:||:|　|:||:|　|:||:|　||::|　|:||::|　|||:
Db　457 A.BC.DA...CSDAN..K.B.CHMMCS.CD.D.ABM.BT...MMHCA.YS..Y......SAB 398

Qy 1245 CGTACAGTCCGTCGTCTTCCCGGAAGTCAATATTACTGCCCCAGTTACTTCTATTGGCAACAC 1304
　　　　　:|:|　|||:|　:|:|：|　:|：||::|：|||：|　:||::|　|||
Db　397 .TBASHSAWY..DTBCAC.NBHSR..MS.AY.T.K.YNC.H.TMY...SHCS.DRTS..C 338

Qy 1305 AGACAGGCCGATTCTTTACAAAACCTCACCACTGGCCATCTACAGGTTCTCCCCATCTACAT 1364
　　　　　::|　|:|　|:||:|　|:||:|　|:||　|::|　|::|　|::|　|:||:|
Db　337 C..Y.AS...NBHBHM...D.MMSBWMYT....MY.C..WKAH.TH..A.C.C.W.S.SYS 278

Qy 1365 ACAGCCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAAAAATCTCAGTCCAAGCCTA 1424
　　　　　:|:|　|::|　|:||:|　|:||:|　||:|　|:||:|　|:||:|　|:||
Db　277 B.M..HDCWT.MYRCMWTD..BH.BCM.YRN.DB.AC.AYT......MA.B.TYYS...AC 218

Qy 1425 TGAGACCCAAGTGAAATTCATCTTTGAGTTGAT 1457
　　　　　:|:|　|:|　|:|　|||　||
Db　217 .....BYCYWS.BAY..HC..Y.HB.T..YMAT 185

RESULT 12
US-09-799-946-5
; Sequence 5, Application US/09799946
; Patent No. US20020099028A1

GENERAL INFORMATION:
  APPLICANT:
    Gudkov, Andrei
    Kazarov, Alexander
    Mazo, Ilya
    Roninson, Igor B
  TITLE OF INVENTION: Methods for Identifying Genetic
    Suppressor Elements and Genes Associated with Malignant
    Growth in Cancer Cells
  NUMBER OF SEQUENCES: 13
  CORRESPONDENCE ADDRESS:
    ADDRESSEE: Allgretti & Witcoff, Ltd.
    STREET: 10 S. Wacker Drive, Suite 3000
    CITY: Chicago
    STATE: Illinois
    COUNTRY: USA
    ZIP: 60606
  COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: PatentIn Release #1.0, Version #1.25
  CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/799,946
    FILING DATE: 06-Mar-2001
    CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
    APPLICATION NUMBER: US/08/204,740
    FILING DATE: 04-MAR-1994
  ATTORNEY/AGENT INFORMATION:
    NAME: No. US200209028Alnan, Kevin E
    REGISTRATION NUMBER: 35,303
    REFERENCE/DOCKET NUMBER: 93,354-C
  TELECOMMUNICATION INFORMATION:
    TELEPHONE: 312-715-1000
    TELEFAX: 312-715-1234
    TELEX: 910-221-5317

INFORMATION FOR SEQ ID NO: 5:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 273 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: cDNA
  SEQUENCE DESCRIPTION: SEQ ID NO: 5:

US-09-799-946-5

Query Match   2.5%;   Score 37.4;   DB 10;   Length 273;
Best Local Similarity   55.8%;   Pred. No. 0.05;
Matches 92;   Conservative   0;   Mismatches   71;   Indels   2;   Gaps   1;

```
Qy  315 CCACCTGTTCACTATCCAAGGCTCGGACCCCCAGCGTTGCAGCCCTACCTGCTCA 374
            ||| ||| ||| |||| ||| |||||||||||| || |||||| | ||| |||
Db  108 CCATCTGCTCGGGCAAGCTAGGCACGGACCCCCAGCGAGAGAAAACCGTGTGCATTTACGGGCA 167

Qy  375 CTTTGATGTGGTGGCCT--GCCCCTGAAGAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTG 432
            ||| || |||| || |  |||||||||| |||| || |||| |||||||||||||||||
Db  168 CCTGGACGTGCCAGCCTGCGGCCCTGGAGGACGGGTGGGACACGCGAGCCCCTTCACCTTGGTG 227

Qy  433 GAGCGTGATGGCGTCGTCATCTATGGTCGGGGGCCACACTGGACGACACAAG 477
            |||| || ||| || | ||||||||||| |||||||||||||||||||| |
Db  228 GAGCGGGAAGGCAAGCTGTATGGGAGAGAGGCTTCCACGGACGATAAG 272
```

  PRIOR APPLICATION NUMBER: US/09/235,076
  PRIOR FILING DATE: 1999-01-20
  NUMBER OF SEQ ID NOS: 38054
  SOFTWARE: FastSEQ for Windows Version 3.0
SEQ ID NO 2716
  LENGTH: 466
  TYPE: DNA
  ORGANISM: Homo sapiens
  FEATURE:
    NAME/KEY: misc_feature
    LOCATION: (1)..(466)
    OTHER INFORMATION: n = A,T,C or G
US-09-918-995-2716

Query Match   2.5%;   Score 37.4;   DB 9;   Length 466;
Best Local Similarity   52.2%;   Pred. No. 0.072;
Matches 108;   Conservative   0;   Mismatches   96;   Indels   3;   Gaps   1;

```
Qy  318 CCTGTTCACTATCCAAGGCTCGGACCCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTT 377
            | |||| ||  ||||| ||| |||||||||||| || |||| | || ||| ||||| |||
Db  250 CGTGTTGACCTGGCCAGGCACCAACCCTACACTCTCCTCCATTCTTGCTCAACTCCCACAC 309

Qy  378 TGATGTGGTGGCCTGCCCCTGAAGAGAAGGCTGGGAGGTGCCCCCATTCTCTGGGTT---GGA 434
            |||||||||||||||  |||||||||| || || ||||||||||||||||||||    |||
Db  310 GGATGTGGTGGCCTGCTTGTTCTTCAAGGAACATTGGAGTCACGACCCCTTTGAGGCCTTCAAGGA 369

Qy  435 GCGTGATGGCGTCATCTATGGTCGGGGGCACACTGACGACAGAACTCTGTGATGGCATT 494
            | ||| ||| || ||||||| |||| |||||||||| || |||| |||| ||||||
Db  370 TTCTGAGGGCTTACATCTATGCCAGGGGTGCCCAGGACGACATGAAGTGCGTCAGCATCCAGTA 429

Qy  495 ACTGCAGGCCTTGGAGCTCCTGCCTGAT 521
            ||| || |||| ||| |||| || |||
Db  430 CCTGGAAGCTGTGAGGAGGCTGAAGGT 456
```

RESULT 14
US-10-108-605-280
  Sequence 280, Application US/10108605
  Patent No. US200201609324A1
GENERAL INFORMATION:
  APPLICANT: Broadus, Julie
  APPLICANT: Stan, Lynn
  APPLICANT: Bachmann, Jane
  APPLICANT: Kandar, Kim
  TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FROM DROSOPHILA MELANOGASTER THAT ENCOD
  TITLE OF INVENTION: PROTEINS ESSENTIAL FOR LARVAL VIABILITY AND USES THEREOF
  FILE REFERENCE: 31113B
  CURRENT APPLICATION NUMBER: US/10/108,605
  CURRENT FILING DATE: 2002-03-27
  PRIOR APPLICATION NUMBER: US 09/761,142
  PRIOR FILING DATE: 2001-01-16
  PRIOR APPLICATION NUMBER: US 60/176,418
  PRIOR FILING DATE: 2000-01-14
  NUMBER OF SEQ ID NOS: 361
  SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 280
  LENGTH: 298
  TYPE: DNA
  ORGANISM: Drosophila melanogaster
US-10-108-605-280

Query Match   2.4%;   Score 36.8;   DB 9;   Length 298;
Best Local Similarity   61.5%;   Pred. No. 0.086;
Matches 59;   Conservative   0;   Mismatches   37;   Indels   0;   Gaps   0;

```
Qy 1177 GCCTTTGACCCCTCCCCCGTTCAGCCCTTCTGATGACAAGGCGTTGGGGTTACCAGCTGCTC 1236
             |||| ||||| ||||||||| ||||||| |||||| ||| |||||| ||||||||||||||||
Db  139 GGCGTGTGTTCTTCTCCGCGCGCGCCGCTTCCCGTTCCGGACCAAGGACGAGGGTCGTCACAGCTGCTC 198

Qy 1237 CGCCAGACCGTACAGTCCGTCTTCCCGGAAGTCAAT 1272
             |||| || ||||| ||| ||||| |||| |||| ||
Db  199 GACCAGGGCCCGGACATCAACACGCGCCAACGTCGAT 234
```

RESULT 13
US-09-918-995-2716
  Sequence 2716, Application US/0918995
  Publication NO. US2003007362JA1
GENERAL INFORMATION:
  APPLICANT: Hyseq, Inc.
  TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
  TITLE OF INVENTION: FROM VARIOUS cDNA LIBRARIES
  FILE REFERENCE: 20411-756
  CURRENT APPLICATION NUMBER: US/09/918,995

RESULT 15
US-10-123-155-298
Sequence 298, Application US/10123155
Publication No. US20030068794A1
GENERAL INFORMATION:
 APPLICANT: Baker,Kevin P.
 APPLICANT: Beresini,Maureen
 APPLICANT: DeForge,Laura
 APPLICANT: Desnoyers,Luc
 APPLICANT: Filvaroff,Ellen
 APPLICANT: Gao,Wei-Qiang
 APPLICANT: Gerritsen,Mary E.
 APPLICANT: Goddard,Audrey
 APPLICANT: Godowski,Paul J.
 APPLICANT: Gurney,Austin L.
 APPLICANT: Sherwood,Steven
 APPLICANT: Smith,Victoria
 APPLICANT: Stewart,Timothy A.
 APPLICANT: Tumas,Daniel
 APPLICANT: Watanabe,Colin K
 APPLICANT: Wood,William
 APPLICANT: Zhang,Zemin
 TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 TITLE OF INVENTION: ACIDS ENCODING THE SAME
 FILE REFERENCE: P3330R1C30
 CURRENT APPLICATION NUMBER: US/10/123,155
 CURRENT FILING DATE: 2002-04-15
 PRIOR APPLICATION removed - See Palm or File Wrapper
 NUMBER OF SEQ ID NOS: 550
 SEQ ID NO 298
  LENGTH: 708
  TYPE: PRT
  ORGANISM: Homo Sapien
US-10-123-155-298

Query Match   2.4%; Score 36.6; DB 9; Length 708;
Best Local Similarity   6.7%; Pred. No. 0.18;
Matches 37; Conservative 142; Mismatches 372; Indels 0; Gaps 0;

Qy   78 GGGCCCGAGAGCGGGGAGCGGGGAGCCATCAAAAGGCGTCCGGAATCCTTCTCAGTTCAGCAAAGA 137
         ::::::
Db   112 VKKMPQLLSVYLEENKLITELPEKCLSELSNLQELYINHNLLSTISPGAFIGLHNLLRLHL 171

Qy   138 GGAACGCGTCGCGATGAAAGAGGCGCCTGAAAGGTGCCATCCAGATTCCAACAGTGACTTT 197
         :: :   ::   ::   ::   ::
Db   172 NSNRLQMINSKWFDALPNLEITLMIGENPIIRIKDMNFKPLINLRSLVIAGINLTEIPDNA 231

Qy   198 TAGCTCTGAGAAGTCCAATACTACAGCCCCTGGCTGAGTTCGGAAAATACATTCATAAAGT 257
         : :   ::   : :   ::
Db   232 LVGLENLESISFYDNRLIKVPHVALQKVVNILKFLDLNKNPINRIRRGDFSNMLHLKELGI 291

Qy   258 CTTTCCTACAGTGGTCAGCAGCACCAGCTTTATCCACCATGAAGTCGTGGAAGAGTATAGCCA 317
         ::   :   :: :   :
Db   292 NNMPELISIDSLAVDNLPDLRIEATNNNPRLSYIHPNAFFRLRPKLESLMLNSNALSALYH 351

Qy   318 CCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCGAGCCCTACCTGCCTGATGGCTCACTT 377
         :   :   :   :   ::   :   :: :
Db   352 GTIESLPNLKEISIHSNPIRCDCVIRWMWMNMNKTNIRFMEPDSLFCVDPPEFQGQNVRQVH 411

Qy   378 TGATGTGGTGGCTGCCCCTGGAAGAAGGCTGGGGAGGTGCCCCCATTCTCTGGGTTGGAGCG 437
         : : ::   ::::   ::   : :   ::   :   ::
Db   412 FRDMMEICLPLIAPESFPPSNLNVEAGSYVSFHCRATAEPQPEIYWITPSGQKLLPNTLTD 471

Qy   438 TGATGGCGTCATCTATGGTCGGGGCACACTGGACGACAAGAAACTCTGTGATGGCATTACT 497
         : :   :   :   ::: ::   :   :
Db   472 KFYVHSEGTLDINGTVTPKEGGLYTCLTATNLVGADLRSVMIKVDGSFPQDNNGSLNIKIRD 531

Qy   498 GCAGGCCTTGGAGCTCCTCGATCAGGAAGAAGTACATCCCCGAAGATCTTTCTTCAITTTC 557
         : : ::   :::::   : : ::::::::   : :   :   :
Db   532 IQANSVLVSWKASSKILKSSVKVKTWTAFVKTENSHAAQSARIPSDVKVYNLTHLNPSTEYKI 591

Qy   558 TCTGGGCCCATGATGAGGAGTCAATCAGGACGAGGGGCTCAGAGGATCTCAGCCCTGCTACA 617
         : ::   : :   : ::::   :   :   :   : :   :
Db   592 CIDIPTIYQRNRKKCVNVTTKGLHPDQREYEKNNTTTLMACLGGLLGIIGVICLLISCLSP 651

Qy   618 GTCAAGGGGCG 628
         :|| ::
Db   652 EMNCDGGHSTV 662

Search completed: July 1, 2003, 09:22:14
Job time : 283 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 10:46:18 ; Search time 86.5 Seconds
                                        (without alignments)
                                        4649.137 Million cell updates/sec

Title: us-10-014-896-1
Perfect score: 2762
Sequence: 1 atggctcagcggtgcgttg..........ctcacctgcacaaactgtga 1509

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Scoring table:
               BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop   6.0 , Fgapext 7.0
               Delop   6.0 , Delext   7.0

Command line parameters:
-MODEL-frame+_n2p.model -DEV-xlp
-O-/cgn2_1/USPTO_spool/US10014896/runat_27062003_104412_10361/app_query.fasta_1.1671
-DB-A_Geneseq_101002 -QFMT-fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT-0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST-45 -DOCALIGN-200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE-LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000000
-USER-US10014896_@CGN_1_114_@runat_27062003_104412_10361 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :
          A_Geneseq_101002:*
    1: /SIDS2/gcgdata/geneseqp-embl/AA1980.DAT:*
    2: /SIDS2/gcgdata/geneseqp-embl/AA1981.DAT:*
    3: /SIDS2/gcgdata/geneseqp-embl/AA1982.DAT:*
    4: /SIDS2/gcgdata/geneseqp-embl/AA1983.DAT:*
    5: /SIDS2/gcgdata/geneseqp-embl/AA1984.DAT:*
    6: /SIDS2/gcgdata/geneseqp-embl/AA1985.DAT:*
    7: /SIDS2/gcgdata/geneseqp-embl/AA1986.DAT:*
    8: /SIDS2/gcgdata/geneseqp-embl/AA1987.DAT:*
    9: /SIDS2/gcgdata/geneseqp-embl/AA1988.DAT:*
   10: /SIDS2/gcgdata/geneseqp-embl/AA1989.DAT:*
   11: /SIDS2/gcgdata/geneseqp-embl/AA1990.DAT:*
   12: /SIDS2/gcgdata/geneseqp-embl/AA1991.DAT:*
   13: /SIDS2/gcgdata/geneseqp-embl/AA1992.DAT:*
   14: /SIDS2/gcgdata/geneseqp-embl/AA1993.DAT:*
   15: /SIDS2/gcgdata/geneseqp-embl/AA1994.DAT:*
   16: /SIDS2/gcgdata/geneseqp-embl/AA1995.DAT:*
   17: /SIDS2/gcgdata/geneseqp-embl/AA1996.DAT:*
   18: /SIDS2/gcgdata/geneseqp-embl/AA1997.DAT:*
   19: /SIDS2/gcgdata/geneseqp-embl/AA1998.DAT:*
   20: /SIDS2/gcgdata/geneseqp-embl/AA1999.DAT:*
   21: /SIDS2/gcgdata/geneseqp-embl/AA2000.DAT:*
   22: /SIDS2/gcgdata/geneseqp-embl/AA2001.DAT:*
   23: /SIDS2/gcgdata/geneseqp-embl/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	2322	84.1	473	23	AAU72909	Human metalloprote
2	827	29.9	166	21	ABB40345	Human ORFX ORF109
3	562.5	20.4	120	22	AAO04998	Human polypeptide
4	472	17.1	116	23	ABP09238	Human ORFX protein
5	451	16.3	93	20	AAY11540	Human 5' EST secre
6	275.5	10.0	441	22	AAG92859	C glutamicum prote
7	274	9.9	422	22	AAB79123	Corynebacterium gl
8	274	9.9	422	22	AAB79819	Corynebacterium gl
9	266.5	9.6	458	22	AAU41311	Propionibacterium
10	253.5	9.2	401	22	ABB62639	Propionibacterium
11	228	8.3	401	22	ABB66170	Drosophila melanog
12	224.5	8.1	408	22	ABB66172	Drosophila melanog
13	217	7.9	340	22	ABB62643	Drosophila melanog
14	207.5	7.5	349	22	ABB62632	Drosophila melanog
15	203	7.3	379	23	ABB47976	Listeria monocytog
16	202	7.3	407	14	AAR30458	Pig aminoacylase I
17	201	7.3	418	23	ABP38707	Staphylococcus epi
18	197.5	7.2	401	22	ABB62394	Drosophila melanog
19	196.5	7.1	318	20	AAY08759	Drosophila melanog
20	195	7.1	391	18	AAW21016	L. monocytogenes d
21	190	6.9	373	23	ABB08102	H. pylori cytoplas
22	190	6.9	388	19	AAW62863	Enzyme similar to
23	190	6.9	429	22	AAG06512	Helicobacter pylor
24	190	6.9	450	21	AAG06511	Arabidopsis thalia
25	189	6.8	474	22	AAG96247	Arabidopsis thalia
26	187.5	6.8	430	21	AAG50167	Putative P. abyssi
27	187.5	6.8	451	21	AAG50166	Arabidopsis thalia
28	185.5	6.7	408	23	ABB08103	Arabidopsis thalia
29	184.5	6.7	408	14	AAR30459	Human peptidase fa
30	182	6.6	458	22	ABP28245	Streptococcus poly
31	175	6.3	331	22	ABP61032	Lactobacillus rham
32	168.5	6.1	246	22	AAB79125	Corynebacterium gl
33	168	6.1	431	23	ABP39493	Staphylococcus epi
34	168	6.1	462	22	ABB66250	Drosophila melanog
35	165	6.0	415	22	ABB66250	Drosophila melanog
36	164.5	6.0	470	23	ABB49465	Listeria monocytog
37	162.5	5.9	460	22	AAU46732	Propionibacterium
38	159.5	5.9	475	22	AAG57236	Amino acid sequenc
39	159.5	5.8	475	22	AAM39882	Amino acid sequenc
40	159.5	5.8	475	23	AAA72910	Human polypeptide
41	155.5	5.6	385	21	AAB42872	Human ORFX ORF2636
42	154.5	5.6	446	23	ABP40280	Staphylococcus epi
43	154.5	5.6	475	22	AAG67143	Amino acid sequenc
44	154.5	5.6	476	22	AAM41668	Human polypeptide
45	152.5	5.5	475	22	AAB93225	Human protein sequ

ALIGNMENTS

RESULT 1
AAU72909
ID   AAU72909 standard; Protein; 473 AA.
XX
AC   AAU72909;
XX
DT   26-FEB-2002 (first entry)
XX
DE   Human metalloprotease partial protein sequence #21.
XX
KW   Human; protease; PCR primer; cytostatic; immunomodulator; cardiant;
KW   vasotropic; antimigraine; analgesic; endocrine; nootropic; tranquiliser;
KW   hypertensive; hypotensive; neuroleptic; neuroprotective; anabolic;
KW   anorectic; antiinflammatory; aspartyl protease; cysteine protease;
KW   metalloprotease; serine protease; cancer; haematopoietic; breast; colon;
KW   lung; prostrate; cervical; brain; ovarian; bladder; kidney; pain;
KW   immune-related disease; cardiovascular disease; neuronal disease;
KW   migraine; sexual dysfunction; mood disorder; attention disorder;
KW   cognition disorder; hypotension; hypertension; psychotic disorder;

```
KW   dyskinesia; metabolic disorder; inflammatory disorder.
OS   Homo sapiens.
XX
PN   WO200183782-A2.
XX
PD   08-NOV-2001.
XX
PF   04-MAY-2001; 2001WO-US14431.
XX
PR   04-MAY-2000; 2000US-201879P.
XX
PA   (SUGE-) SUGEN INC.
XX
PI   Plowman GD, Whyte D, Sudarsanam S, Manning G, Caenepeel S;
PI   Payne V;
XX
DR   WPI; 2002-041502/05.
DR   N-PSDB; AAS97192.
XX
PT   Novel protease polypeptide useful for screening for substances that may
PT   be used to treat, e.g., cancers, immune-related diseases,
PT   cardiovascular disease, migraine, pain, psychotic and inflammatory
PT   disorders -
XX
PS   Claim 28; Figure 21; 232pp; English.
XX
CC   The invention relates to an isolated, enriched, or purified protease
CC   polypeptide (II) and polynucleotide (II) encoding (I). (I) may be used to
CC   screen for substances (S) that may modulate its activity. Administering
CC   s (which modulates protease activity in vitro) may be used to treat a
CC   disease or disorder selected from cancers (e.g., of tissues, of blood or
CC   haematopoietic origin, of the breast, colon, lung, prostrate, cervical,
CC   brain, ovarian, bladder or kidney), immune-related diseases and
CC   disorders, cardiovascular disease, brain or neuronal-associated diseases
CC   (e.g., central or peripheral nervous system diseases, migraine, pain,
CC   sexual dysfunction, hypotension, mood disorders, attention disorders, cognition
CC   disorders, hypertension, psychotic disorders, neurological
CC   disorders and dyskinesias), metabolic disorders and inflammatory
CC   disorders. (I) may also be useful as a diagnostic tool for a disease or
CC   disorder such as those above. AAU72876-AAU72910 represent human
CC   protease amino acid sequences of the invention.
XX
SQ   Sequence 473 AA;
```

Alignment Scores:

Pred. No.:	2.18e-215	Length:	473
Score:	2322.00	Matches:	470
Percent Similarity:	93.45%	Conservative:	1
Best Local Similarity:	93.25%	Mismatches:	2
Query Match:	84.07%	Indels:	31
DB:	23	Gaps:	2

```
US-10-014-896-1 (1-1509) x AAU72909 (1-473)

Qy    1 ATGGCTCAGCGGTGCGGTTGCCGTTGCCTGGCTATGCTGCTCCCATAGTTTTCCCT   60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MetAlaGlnArgCysValCysValLeuAlaLeuAlaLeuMetLeuLeuValPhePro   20

Qy   61 ACCGTCTCCAGAtcGAATGGGCCCGGAGCAGCGGGAGCATCAAAGGGCGTCCGAAtCCCT  120
        ||||||||||||::|||||||||||||||||||||||||||||||||||||||||||
Db   21 ThrValSerArgSerMetGlyProArgSerGlyValTyrGlnArgAlaSerArgIlePro   40

Qy  121 TCTCAGTTCAGCAAAAGAAGACGCGTGCGCGATGAAAAGAAGGTGCCATCCAG        180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerGlnPheSerLysArgGlyValAlaMetLysGlnLeuAlaLeuLeuLysGlyValAlaIleGln   60

Qy  181 ATTCCACACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTTGTGAGTTCGGA  240
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly   80

Qy  241 AAATACAATTCATAAAAGTCTTTCCTACAGTGGTCAGCAGCCAGCACCAGCATGAAGTC   300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 LysThrTyrIleArgLysValPheProThrValValSerThrSerPheIleGlnHisGlyVal  100

Qy  301 GTGGAAGAGTATAGCCACCACCTGTTCACTATCCAAGGGCCCAGCTTGCAGCCCTAC     360
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  101 ValGluGluTyrSerHisLeuPheThrIleGlnGlyXaaSerXaaSerXaaSerAspProSerLeuGlnProTyr  120

Qy  361 CTGGCTGATGGCTCACTTTGATGTGGTGCGCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCA   420
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 LeuLeuLeuMetAlaHisPheAspValValProAlaProGluGluGlyTrpGluValProPro  140

Qy  421 TTCTCTGGGTTGGAGCGTGGCGTGATCGTCATCATGGGTCGGGGACACACTGGACGACAAGAAC  480
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  141 PheSerGlyLeuGluArgGlyAspGlyValValIleValGlyTyrGluValGlyAspArgAsn  160

Qy  481 TCTGTGGATGGCAATACTACCGAGGCCCTTGGAGCTCCCTGCTGATCAGGAGAGTACATCCCCGA  540
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  161 SerValMetAlaLeuLeuGlnLeuLeuAsnGluLeuLeuIleArgLysTyrIleProArg  180

Qy  541 AGATCTTTCTTCATTTCTCTGGGCCACAGATGAGGAGGTCATCAGGAGCAGGGGCTCAGAGG   600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 ArgSerProPhePheIleSerLeuGlyHisAspGluGluSerGlyThrGlyAlaGlnArg  200

Qy  601 ATCTCAGCCCCTGCCTACAGTCAAGGGGGTCCAGCTAGCCTTCATTGTGGACGAGGGGGGC    660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  201 IleLeuSerAlaLeuLeuGlnSerArgGlyValGlnLeuLeuAlaPheIleValAlaPgGluGlyGly  220

Qy  661 TTCAGCTTGGATGAGTTCATTCCTAACTTCAAGAAGCCCATGGCTCCTGATCAGTCTCA     720
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  221 PheIleLeuLeuAspAsnPheIleProAsnPheLysLysProIleAlaLeuIleAlaValSer  240

Qy  721 GAGAAGGGGTTCCATGAACCTCACGCTGCAAGTAAAACAGACTTCCAGGCCACTCTTCAGCT   780
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GluLysGlyGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla  260

Qy  781 CCTCCAAAGGAGGACAAAGCATTGGCATCCTTGCAGCTGCCTGCACCGCCGATTGGAGAGACA  840
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  261 ProProLysGluThrSerIleGlyTyrIleLeuGlnValAlaAlaAlaValSerArgLeuGlnThr  280

Qy  841 CCAAATGCCTATCATATTTGGGAAGCGGGGACAGTGGTCGAACAAGCAACTGGCAAAT      900
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  281 ProMetProIleIlePheGlySerGlyThrValValThrValLeuGlyLeuLeuAlaAsn  300

Qy  901 GAGTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTAATTGAACCACACTTATA   960
        -----------------------------------------------------------
Db  300 -----------------------------------------------------------  300

Qy  961 AGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCCACTC 1020
        |||||||||||||||||||||
Db  301 Glu--ValTyrGlyIleGlyLysSerLeuAsnGlnCysAsnAsnGlnAspHisHisGlyThrH  320

Qy 1021 ACCATAATTCAAAGCAGGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCAGGCCACAGTC 1080
        |||||||||||||| .||
Db  320 isHisIleGlnSerArg-Val-----------------------AlaGlnAlaThrVal  331

Qy 1081 AACTTCCGGATTCACCCTGGACAGAGTCCAAGAGGTCCTAGAAACTCACGAAGAACATT  1140
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  332 AsnPheArgIleHisProGlyGlnSerProGlyValLeuGluThrHisGluGluHisIle  351

Qy 1141 GTTGGCTGATAACACAGAGTCCACAGTGTTGAGTGGCCTTTGACCCCGTCCCCGTCAGC  1200
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  352 ValAlaAspAsnThrGluSerThrValLeuSerAlaPheAspProLeuProValSer  371

Qy 1201 CCTTCTCGATGACAAGGCCTTGGGCTACCACAGTCCTCCGCCACGACTCCGTTCTTC    1260
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  372 ProSerArgAspLysAlaLeuGlyTyrGlnSerProProAlaThrThrProPhePhe  391

Qy 1261 CCGGAAGTCAAATATTACTGCCCCAGTTACTTCTCTATTGGCAACAGCAGACAGGAGATTCTTT 1320
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  392 ProGluValLysAsnIleThrAlaProValThrSerIleGlyAsnThrAspSerSerArgPhePhe  411

Qy 1321 ACAAAACCTCACCACTGGAGAATCACACAGGTTCTACCCCATCTACAATACAGCCGAAGACTTC 1380
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  412 ThrAsnLeuThrThrGlyIleThrTyrArgPheTyrProIleTyrIleGlnProGluAspPhe  431
```

```
Qy    1381 AAACGCATCCATGGAGTCAACGAGAGAAAATCTCACTCACCAAGACCCTATGAGACCCAAGTGAAA 1440
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     432 LysArgIleHisGlyValAsnGluLysIleSerValGlnAlaTyrGluThrGlnValLys 451

Qy    1441 TTCATCTTTGAGTTGATTGAATTCAGAATGTCGACACAGACCAGGAGCCCAGTTTCTCACCTGCAC 1500
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     452 PheIlePheGluLeuIleGlnAsnAlaAspThrAspGlnGluProValSerHisLeuHis 471

Qy    1501 AAACTG 1506
             ||||||
Db     472 LysLeu 473
```

```
RESULT 2
ID   AAB40345
XX
AC   AAB40345;
XX
DT   08-FEB-2001 (first entry)
XX
DE   Human ORFX ORF109 polypeptide sequence SEQ ID NO:218.
XX
KW   Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW   vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW   anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW   immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW   hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW   antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW   antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW   neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW   cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW   cholesterol ester storage; systemic lupus erythematosus; infection;
KW   severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW   allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW   bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW   thrombosis; contraceptive.
XX
OS   Homo sapiens.
XX
KW
XX
PN   WO200058473-A2.
XX
PD   05-OCT-2000.
XX
PF   31-MAR-2000; 2000WO-US08621.
XX
PR   31-MAR-1999; 99US-0127607.
PR   02-APR-1999; 99US-0127636.
PR   05-APR-1999; 99US-0127728.
PR   30-MAR-2000; 2000US-0540763.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shimkets RA, Leach M;
XX
DR   WPI; 2000-603362/57.
DR   N-PSDB; AAC74554.
XX
PT   Novel nucleic acids and peptides derived from open reading frame X,
PT   useful for treating e.g. cancers, proliferative disorders,
PT   neurodegenerative disorders and cardiovascular disease -
XX
PS   Claim 11; Page 547; 5507pp; English.
XX
CC   AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC   which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC   sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC   antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
CC   anticonvulsant; osteopathic; antiarthritic; immunosuppressant;
CC   immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC   antidiabetic; hypotensive; dermatological; immunosuppressive;
CC   antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC   antithyroid; and antianaemic. The sequences can be used for determining
```

```
CC   the presence of or predisposition to, or preventing or treating
CC   pathological conditions associated with an ORFX-associated disorder. The
CC   nucleic acids can be used to express ORFX proteins in gene therapy
CC   vectors. The proteins and nucleic acids may be used to treat cancers,
CC   proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC   graft vs host disease, cardiovascular disease, diabetes mellitus,
CC   hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
CC   erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC   bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC   allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC   nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC   coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ   Sequence  166 AA;
```

```
Alignment Scores:
Pred. No.:
Score:                        4.41e-71
Percent Similarity:           827.00            Length:          166
Best Local Similarity:        99.39%            Matches:         164
Query Match:                  99.39%            Conservative:      1
                              29.94%            Mismatches:        1
                              21                Indels:            0
                                                Gaps:              0
DB:

US-10-014-896-1 (1-1509) x AAB40345 (1-166)

Qy   142 CGCGTGCGGATGAAAAGAGGGCGCTGAAAAGGTGCCCATCCAGATTCCAACAGTGACTTTAGC 201
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     2 ArgValAlaMetLysGluAlaLeuLysGlyAlaLeuLysGlyAlaIleGlnIleProThrValThrPheSer 21

Qy   202 TCTGAGAAGTCCAATACTACAGCCCCTGGCCTGAGTTCGGAAAATACATTCATAAAAGTCTTT 261
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    22 SerGluLysSerAsnThrThrAlaLeuAlaGluProGlyLeuSerSerGluAsnThrPhePhe 41

Qy   262 CCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAAGTCCTGGAAGAGTAAAGCCACCTG 321
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    42 ProThrValSerThrSerPheIleGlnHisGluValProGluGluLeuTyrSerHisLeu 61

Qy   322 TTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCCTGCTGATGGGTCACTTTGAT 381
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 PheThrIleGlnGlySerAspProSerLeuGlnProTyrLeuLeuMetAlaHisPheAsp 81

Qy   382 GTGGTGCCTGCCCTGAAAGAAGGGCTGGGAGGTTGCCCCCATTCTCTGGGTTGGAGGTGAT 441
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    82 ValValProAlaProGlyLysGlyTrpGluValProPheSerGlyLeuGluArgAsp 101

Qy   442 GGCGTCATCTATGGTCGGGGACACCTGGACAAGCAGAACTCTGTGATGGCATTACTGCAG 501
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   102 GlyValIleTyrGlyArgGlyThrLeuAspLysGlnAsnSerValMetAlaLeuLeuGln 121

Qy   502 GCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGGAAGAGTCTTTCTTCATTTTCTCTG 561
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   122 AlaLeuGluLeuLeuLeuIleArgLysTyrIleProArgArgSerPhePheIleSerLeu 141

Qy   562 GGCCATGATGAGGAGTCATCAGGGACAGAGGGGCTCAGAGGAGTCCACCCCTGCTACAGTCA 621
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   142 GlyHisAspGluGluSerSerGlyThrGlyAlaGlnArgIleSerAlaLeuLeuGlnSer 161

Qy   622 AGGGGCGTCCACGTA 636
             |||||||||||||||
Db   162 ArgGlyValGlnLeu 166
```

```
RESULT 3
ID   AAO04998
XX
AC   AAO04998;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human polypeptide SEQ ID NO 18890.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
```

```
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
KW   nervous system disorders; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO20164835-A2.
XX
PD   07-SEP-2001.
XX
PF   26-FEB-2001; 2001WO-US04927.
XX
PR   28-FEB-2000; 2000US-0515126.
PR   18-MAY-2000; 2000US-0577409.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT;
XX
DR   WPI; 2001-514838/56.
DR   N-PSDB; AAI84929.
XX
PT   Isolated nucleic acids and polypeptides, useful for preventing
PT   diagnosing and treating e.g. leukaemia, inflammation and immune
PT   disorders -
XX
PS   Claim 20; SEQ ID NO 18890; 1399pp + Sequence Listing; English.
XX
CC   The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC   the encoded proteins (AAO00010-AAO13910), that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation.
CC   Note:The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 120 AA;

Alignment Scores:
Pred. No.:              1.39e-45
Score:                  562.50        Length:          120
Percent Similarity:     79.72%        Matches:         114
Best Local Similarity:  79.72%        Conservative:    0
Query Match:            20.37%        Mismatches:      0
DB:                     22            Indels:          29
                                      Gaps:            1

US-10-014-896-1 (1-1509) x AA004998 (1-120)

Qy          1 ATGGCTCAGGTGCGGGTGCGTTTGCGTGCGTGGCCCTGGTGCTGCTCCTAGTTTTCCCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          7 MetAlaGlnArgCysValCysValLeuLeuAlaLeuValAlaLeuValAlaLeuMetLeuLeuLeuValPhePro 26

Qy         61 ACCGTCTCTCCAGATCGAATGGGCCCGAGGAGGCGGGAGCANTCAAAGGGCGTTCGCGGAATCCCT 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         27 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro 46

Qy        121 TCTCAGTTCAGCAAAGAGGAACGCGTCGCGATGAAAGAGGCGCGTGAAAGGTGCCATCCAG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         47 SerGlnPheSerLysGluGlnArgValAlaMetLysGluAlaLeu-------------- 61

Qy        181 ATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACACGCCTGGCTGAGTTCGGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 -------------------------------------------------- 61

Qy        241 AAATACATTCATAAAAGTCTTTCCTACAGTGGTCAGCACCAGTGTTATCCAGCATGAAGTC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         62 --------------LysValPheProThrValSerThrSerPheIleGlnHisGluVal 77
```

```
Qy        301 GTGGAAGAGTATAGCCACCTGTTCACTATCCCAAGGCTCGGACCCGACCCTRAC 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         78 ValGluGluTyrSerHisLeuPheThrIleGlySerAspProSerLeuGlnProTyr 97

Qy        361 CTGCTGAATGCGCTCACTTTGATGTGGTGTGGTGCCTGCGCCCCTGAAGAAGGCTGGGAAGGGTGCCCCCA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         98 LeuLeuMetAlaHisPheAspValValProGlyAlaAlaProGluGluGlyTrpGluValProPro 117

Qy        421 TTCTCTGGG 429
              |||||||||
Db        118 PheSerGly 120

RESULT 4
ABP09238
ID   ABP09238 standard; Protein; 116 AA.
XX
AC   ABP09238;
XX
DT   25-JUN-2002 (first entry)
XX
DE   Human ORFX protein sequence SEQ ID NO:18458.
XX
OS   Homo sapiens.
XX
KW   Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW   hyperproliferative disorder; Psoriasis; benign tumour; haemorrhage;
KW   degenerative disorder; osteoarthritis; diabetes mellitus; neurodegenerative disorder;
KW   cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW   hypertension; hypothyroidism; cholesterol ester storage disease;
KW   immune deficiency; immune disorder; infectious disease;
KW   autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW   myasthenia gravis.
XX
PN   WO20192523-A2.
XX
PD   06-DEC-2001.
XX
PF   29-MAY-2001; 2001WO-US10836.
XX
PR   30-MAY-2000; 2000US-206132P.
PR   29-AUG-2000; 2000US-228716P.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Shinkets RA, Leach MD;
XX
DR   WPI; 2002-106308/14.
DR   N-PSDB; ABN24990.
XX
PT   Novel human polypeptides and polynucleotides useful for diagnosing,
PT   preventing and treating cardiovascular disease, neurodegenerative,
PT   hyperproliferative disorders and autoimmune disorders -
XX
PS   Disclosure; SEQ ID 18458; 1037pp; English.
XX
CC   The present invention describes substantially purified human proteins
CC   (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
CC   in the specification). ABN15762 to ABN27252 encode the human ORFX
CC   proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC   treating or preventing a pathology associated with an ORFX-associated
CC   disorder in humans, and in the manufacture of a medicament for treating a
CC   syndrome associated with ORFX-associated disorder. ORFX polynucleotide
CC   sequences can be used in gene therapy. ORFX sequences can be used in the
CC   treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
CC   psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
CC   osteoarthritis, cardiovascular diseases, diabetes mellitus, systemic
CC   lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
CC   storage disease, various immune deficiencies and disorders, infectious
CC   diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
CC   arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
CC   disease and autoimmune inflammatory eye disease. ORFX proteins are also
```

CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
CC bone degenerative disorders, or periodontal disease, and for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues and conditions resulting from
CC systemic cytokine damage.
CC N.B. The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 116 AA;

Alignment Scores:
Pred. No.:
Score: 7.68e-37
Matches: 472.00
Percent Similarity: 90.43%
Best Local Similarity: 77.39%
Query Match: 77.09%
DB:

US-10-014-896-1 (1-1509) x ABP09238 (1-116)

Length: 116
Matches: 89
Conservative: 15
Mismatches: 11
Indels: 0
Gaps: 0

Qy 178 CAGATTCCAACGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTC 237
       ::|||||||||||| ||| :::||||||||||||||||||||||||||||||||||||||
Db 2 ArgIleProThrValSerPheSerHisGluGluSerAsnThrThrAlaLeuAlaGluPhe 21

Qy 238 GGAAAAACAATTCATTAAAAGTCTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAA 297
       |||:|||||||||| |||| ||| ::||||||||||||||||||||||||||||||||||
Db 22 GlyGluTyrIleArgLysAlaPheProThrValPheHisSerAsnLeuIleGlnHisGlu 41

Qy 298 GTCGTGGAAAGAGTATAGCCACCCTGTTCACTATACCAAGGCTGGGACCCCAGCTTGCAGCCC 357
       |||||| ::|||||||||||||||||||||||:::||||||||||||||||||||||||||
Db 42 ValValGlyLysTyrSerHisLeuPheThrValArgGlySerAspProSerLeuGlnPro 61

Qy 358 TACCTGCTGATGGCTCACTTTGATGTGGTGCCCCCTGAAGAAGGCTGGGAGGGTGCCC 417
       ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db 62 TyrMetLeuMetAlaHisPheAspValValProAlaSerGluGluGlyTrpGluValPro 81

Qy 418 CCATTCTCTGGGTTGGGAGCGTGATGGGCTCATCTATAGGTCGGGGCCACACTGGACGACAAG 477
       |||||||||||| || |||||||||||||||||||||||||||||||||||||||||:
Db 82 ProPheSerGlyLeuGlyAlaGlyLysProSerLysSerHisTyrGlyAlaLeuAlaAspAsnLys 101

Qy 478 AACTCTGTGATGGCAATTACTGCAGGGCCTTGGAGGCTCCTGCTGATC 522
       |||||||||||||| :::||||||||||||||||||||||||||||||
Db 102 AsnSerValMetAlaValLeuGlnAlaLeuGluLeuLeuLeuIle 116

RESULT 5
AAY11540
ID AAY11540 standard; Protein; 93 AA.
AC AAY11540;
XX
DT 16-JUN-1999 (first entry)
DE Human 5' EST secreted protein SEQ ID NO:192.
XX
KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
KW forensic; gene therapy; chromosome mapping; signal peptide;
KW upstream regulatory sequence; cytokine activity; cell proliferation;
KW differentiation; haematopoiesis regulation; tissue growth regulation;
KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
KW thrombolytic; anti-inflammatory; tumour inhibition.
XX
OS Homo sapiens.
XX
PN WO906439-A2.
PD 11-FEB-1999.
XX
PF 31-JUL-1998; 98WO-IB01233.
PR 01-AUG-1997; 97US-0904468.
XX
PA (GEST) GENSET.

XX
PI Duclert A, Dumas Milne Edwards J, Lacroix B;
XX
DR WPI; 1999-153700/13.
DR N-PSDB; AAX40258.
XX
PT New nucleic acids encoding human secreted proteins - obtained from
PT cDNA libraries derived from liver, lung, large intestine, colon,
PT thyroid and pancreas tissue
XX
PS Claim 27; Page 312; 398pp; English.
XX
CC AAX40251 to AAX40397 represent 5' expressed sequence tags (ESTs) for
CC human secreted proteins, and encode the proteins given in AAY11533 to
CC AAY11679, respectively. The proteins given represent the signal peptide
CC and an N-terminal fragment of a secreted protein. The nucleic acid
CC sequences can be used for producing secreted human gene products. They
CC can also be used to develop products for diagnosis and therapy. The
CC proteins obtained may have cytokine activity, cell
CC proliferation/differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptide can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.
XX
SQ Sequence 93 AA;

Alignment Scores:
Pred. No.:
Score: 7.3e-35
Matches: 451.00
Percent Similarity: 98.94%
Best Local Similarity: 98.94%
Query Match: 16.33%
DB:

US-10-014-896-1 (1-1509) x AAY11540 (1-93)

Length: 93
Matches: 93
Conservative: 0
Mismatches: 0
Indels: 1
Gaps: 0

Qy 1 ATGGCTCAGCCGGTGCGTTTGCCTGCTGCCCTGGTCGCTGCTAGCTGCTCCCAGTTTTCCCT 60
      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1 MetAlaGlnAsnAlaArgCysValCysValLeuLeuAlaLeuAlaMetLeuLeuValPhePro 20

Qy 61 ACCGTCTCCAGTATCGATAGGGCCCGAGGAGCGGGGGAGCATCAAAGGGGGTCTCGGGAATCCCT 120
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 21 ThrValSerArgSerMetGlyProArgSerGlyTyrGluHisGlnArgGlnAlaSerArgIlePro 40

Qy 121 TCTCAGTTCAGCAAAAGAGGACGGCGTGCGGAACAGACAGGCCGGCTGAAAAGGTGCCCATCCAG 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 41 SerGlnPheSerLysGlyLeuArgValAlaMetLysGluAlaLeuArgLysGlyAlaIleGln 60

Qy 181 ATTCCAACAGTTGACTTTTAGCTCTGAGAAGTCCAATACTACACAGCCCTGGCTGAGTTCGGA 240
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 61 IleProThrValThrPheSerSerGluLysSerAsnThrThrAlaLeuAlaGluPheGly 80

Qy 241 AAATACAATTCATTAAAAGTCTTTCCTACAGTGGTCAGCACCA 280
       ||||||||||||||||||||||||||||||||||||||||||||
Db 81 -AsnThrPheIleLysSerPheLeuGlnTrpSerAlaPro 93

RESULT 6
AAG92859
ID AAG92859 standard; Protein; 441 AA.
AC AAG92859;
XX
DT 26-SEP-2001 (first entry)
DE C glutamicum protein fragment SEQ ID NO: 6613.
XX
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;

KW   organic acid synthesis.
XX
OS   Corynebacterium glutamicum.
XX
PN   EP1108790-A2.
XX
PD   20-JUN-2001.
XX
PF   18-DEC-2000; 2000EP-0127688.
XX
PR   16-DEC-1999; 99JP-0377484.
PR   07-APR-2000; 2000JP-0159162.
PR   03-AUG-2000; 2000JP-0280988.
XX
PA   (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI   Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI   Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR   WPI; 2001-376931/40.
DR   N-PSDB; AAH68078.
XX
CC   Novel polynucleotides derived from Coryneform bacteria, for identifying
CC   mutation point of a gene, measuring expression of a gene, analysing
CC   expression profile or pattern of a gene and identifying homologous gene
PN   -
PS   Claim 17; SEQ ID NO: 6613; 246pp + Sequence listing; English.
XX
CC   The present invention provides a number of nucleotide and protein
CC   sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC   sequences are useful for identifying the mutation point of a gene derived from a
CC   mutant of coryneform bacterium, measuring expression amount and
CC   analysing the expression profile or pattern of a gene derived
CC   from Coryneform bacterium, and identifying a homologue of a gene derived
CC   from coryneform bacterium. Coryneform bacteria are useful for producing
CC   amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC   particularly L-lysine. The present sequence is a protein described
CC   in the exemplification of the invention.
CC   Note:The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from the
CC   European Patent Office.
XX
SQ   Sequence   441 AA;

Db   80 HisThrAspValValProValaValaSerProPheGlyAlaGlu 99

Qy   433 GAGCGTGATGGCGTCGTCATCCTATGGGTCGGGGACACTGGGACACAAGAGAACTCTGTGATGGCA 492
          ||| ||| :::||| :::|| || ||| :::||| || :::||| |||
Db   100 IleSerAspGlyGlnIleThrTrpGlyArgGlySerValAsp------------MetLeu 114

Qy   493 TTACTGCAGGGCCTTGGAGCTCCTGCTGATCAGGAGTACAITCCCGAAGATCT---- 546
          |||||| :::||| |||:::||| :::||| :::||| ||| ||
Db   115 PheIleThrAlaThrGlnAlaLeuValThrArgGln---ValAlaAargGluGlyLeu 133

Qy   547 --TTCTCATTTCTCTGGGCCATGAATGAGAGGGCT 594
          ||:::||| :::||| |||:::||| |||
Db   134 ArgGlyThrLeuThrPheValGlyValAlaAspGluGluAlaArgGlyGlyLeuGlyAla 153

Qy   595 CAGAGGAATCTCA-------- 621
          |||:::|||
Db   154 LysTrpLeuSerGluGluHisGlnLeuPheSerTrpLysAsnCysLeuSerGluSer 173

Qy   622 AGGGGCGGTCCAGCTACCCTTCATTGTGGACGAGGGGCTTCATCTTGGATGATTTCATT 681
          || ||| ||| :::||| ||| :::|||
Db   174 GlyGlySerHisLeuPro------ValHisAspGly 183

Qy   682 CCTAACTTCAAGAAGCCCATGCACTTCAGAGAAGGGGTTCCATGAACCTC 741
          :::||| |||:::||| ||||||
Db   184 ------SerAspAlaValIleAsnValGlyGluLysAlaAlaGlnArg 199

Qy   742 ATGCTGCAAGTAAAAACATGACTTCAGGCCACTCTTCAGCTCCT---CCAAAGGAGGACAGC 798
          ||| :::||| |||
Db   200 ArgIleHisValAsnGlyAspAlaGlyHisGlySerIleProPheAspAspAsnSerAla 219

Qy   799 ATTGGCCATCCTTGCAGCTGCTGTTCAGCCGATTGGAGGACAGCCAAATGCCTATCATA- 855
          ||| :::||| :::||| |||:::|||
Db   220 IleValLysIleGlyGlyValAlaAlaArgIleGlnAlaIleAlaAlaAspLeuLysValAlaAlaLys 239

Qy   856 ------TTTGGAAGCGGGGACAGTGGTGACTGTATTGCAGCAACTGCAAATGAGTTT-- 906
          ||| :::|||
Db   240 AspAspAlaProIleArgThrGlnGlyPheVal------------GlnAlaHisArgPheAsp 253

Qy   907 CCCTTCCCTGTCAATATAAACCCTGAGC------AACCCATGGCTGTATTGAACCACTT-- 957
          ||| :::||| :::|||
Db   254 ProGluThrGluGlnAlaLeuLeuSerGlyThrSerProGluAlaThrYrAlaGluPheGly 273

Qy   958 ---ATAAGCAGGTTTATTGAGGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACG 1014
          :::|||||||||
Db   274 GlyLeuSerArgPhe------AlaHisAlaValSerHisHisLeuThrIle 287

Qy   1015 GCACTCACCATATTCAAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCC 1074
          || ||| :::||| :::||| ||||||:::|||
Db   288 AlaGlnThrValValAlaArgAlaGlyGlnAlaAlaIleAsnValLeuProSerHisAlaTyrLeu 307

Qy   1075 ACAGTCAACTTCCCGGATTCACCGGCAACA---------GTCCAAGAGGTCCTAGAA 1125
          ||| :::|||
Db   308 GluLeuAspIleArgThrLeuProGlyThrThrAsnAspThrYrValAspAspThrLeuArg 327

Qy   1126 CTCACGGAAGACACAATTGTGGGTGGATAGAACGAGTCCAGTTGCCAGTGGTGTTGAGTGGCC 1185
          ||| :::||| |||:::|||
Db   328 AlaAlaLeuGlyAspLeuAlaAlaGlyValGlValGluGlnLeuHisLeuIleSer------Glu 345

Qy   1186 CCCCTCCCCGTCAGCCCTTCTTGATGACAAGGGCCTTGGGGTACCAGCTCGCCGCCAGACC 1245
          || ||| :::||| |||
Db   346 GluLysAlaThrValSerProThrArgAspSerArgLeu------TyrArgAsnThrLeuLysVal 363

Qy   1246 GTACAGTCCGTCTCCCCCGAAGTCAATAATTACGTCCCCCAGTTACTTCTTATTGGCAACA 1305
          :::||| |||||| :::|||
Db   364 LeuGlyAspPhePheProAspAlaProVal---ValProIleIleSerSerGlyGlySer 382

Qy   1306 GACAGCCGAGTTCTTTACAAAACCTCACCACTGGCATCACAGGAGTTCTACCCATCTACATA 1365
          ||| ||||||| :::||| |||:::|||
Db   383 AspLeuAspPheGlyArgArgLeu---GlyValGlyTyrGlyPheAlaValHisAla 401

Qy   1366 CAG------ 1404
          :::|||
Db   402 ArgGluArgThrLeuAlaGluAlaMetGlyGlnLeuHisSerHisAspGluAlaLeuThrTyr 421

Qy   1373 CACTTTGATGTGGTGCCTCGCCCCTGAAGAAGAGGGTGGGAGGTGCCCCCATTCTCTGGGTTG 1432
          ||| |||||||||||||||||||

```
Qy     1405  -----------AAAATCTCAGTCCCAAGCCTAYGAGAGACCCAAGTGAA  1440
                         ::::::::::|||::    ||||::::::  ||||::::
Db      422  LeuGluAspLeuGluLeuThrValArgGlyTyrAspSerValAlaArg  437
```

Db

Qy

DT   30-APR-2001 (first entry)

XX

DE   Corynebacterium glutamicum HA protein sequence SEQ ID NO:202.

XX

PR   Corynebacterium glutamicum; homeostasis; adaptation; HA protein;
PR   fine chemical production; organic acid; proteinogenic amino acid or
PR   nonproteinogenic amino acid; purine base; pyrimidine base; nucleoside;
PR   nucleotide; lipid; saturated fatty acid; unsaturated fatty acid; diol;
PR   carbohydrate; aromatic compound; vitamin; cofactor; polyketide; enzyme;
PR   diagnosis; Corynebacterium diphtheriae; genetic engineering;
PR   Brevibacterium; environmental condition.

XX

OS   Corynebacterium glutamicum.

XX

PN   WO200100842-A2.

XX

PD   04-JAN-2001.

XX

PF   23-JUN-2000; 2000WO-IB00911.

XX

PR   25-JUN-1999; 99US-0141031.
PR   08-JUL-1999; 99DE-1031636.
PR   09-JUL-1999; 99DE-1032125.
PR   09-JUL-1999; 99DE-1032126.
PR   09-JUL-1999; 99DE-1032127.
PR   09-JUL-1999; 99DE-1032128.
PR   09-JUL-1999; 99DE-1032129.
PR   09-JUL-1999; 99DE-1032226.
PR   14-JUL-1999; 99DE-1032920.
PR   14-JUL-1999; 99DE-1032922.
PR   14-JUL-1999; 99DE-1032924.
PR   14-JUL-1999; 99DE-1032928.
PR   14-JUL-1999; 99DE-1032930.
PR   14-JUL-1999; 99DE-1032933.
PR   14-JUL-1999; 99DE-1032935.
PR   14-JUL-1999; 99DE-1032973.
PR   14-JUL-1999; 99DE-1033002.
PR   14-JUL-1999; 99DE-1033003.
PR   14-JUL-1999; 99DE-1033005.
PR   14-JUL-1999; 99DE-1033006.
PR   31-AUG-1999; 99DE-1041378.
PR   31-AUG-1999; 99DE-1041379.
PR   31-AUG-1999; 99DE-1041390.
PR   31-AUG-1999; 99DE-1041391.
PR   03-SEP-1999; 99DE-1042088.

XX

PA   (BADI) BASF AG.

XX

PI   Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;

XX

DR   WPI; 2001-061974/07.
DR   N-PSDB; AAF71238.

XX

PT   New Isolated Corynebacterium glutamicum nucleic acid for production or
PT   modulation of production of fine chemicals such as amino acids,
PT   nucleosides, nucleotides, lipids, fatty acids, carbohydrates, vitamins
PT   or enzymes -

XX

PS   Claim 20; Page 413-414; 712pp; English.

XX

CC   AAF71138 to AAF71357 encode the Corynebacterium glutamicum homeostasis
CC   and adaptation (HA) proteins given in AAB79023 to AAB79242. The

CC   C. glutamicum HA genes (I) can be used in vectors for expression in host
CC   cells and production of fine chemicals, such as, an organic acid,
CC   proteinogenic or nonproteinogenic amino acid (preferred), purine or
CC   pyrimidine base, nucleoside, nucleotide, lipid, saturated or unsaturated
CC   fatty acid, diol, carbohydrate, aromatic compound, vitamin, cofactor,
CC   polyketide or enzyme. The amino acids produced can be lysine, glutamine,
CC   glutamate, alanine, aspartate, glycine, serine, threonine, methionine,
CC   cysteine, valine, leucine, isoleucine, arginine, proline, histidine,
CC   tyrosine, phenylalanine, or tryptophan. The presence of (I) or HA proteins
CC   diphtheriae. (II) can be used to map the C. glutamicum genome or can be
CC   used as markers for genetically engineered Corynebacterium or
CC   Brevibacterium. The HA proteins encoded by the (II) are used to maintain
CC   homeostasis in C. glutamicum or help the microorganism to adapt to
CC   be modulated. The proteins encoded by (I) or HA proteins can be
CC   used for diagnosing the presence or activity of Corynebacterium
CC   different environmental conditions.

XX

SQ   Sequence   422 AA;

Alignment Scores:
Pred. No.:
Score: 274.00   Matches: 106
Percent Similarity: 42.55%   Conservative: 71
Best Local Similarity: 25.48%   Mismatches: 161
Query Match: 9.92%   Indels: 78
DB: 22   Gaps: 20

Pred. No.:   2.16e-17   Length: 422

US-10-014-896-1 (1-1509) x AAB79123 (1-422)
```

```
Qy    313  AGCCACCTGTTCACTATCCAAGGCTTCGGACACCCCAGAGCTTGCAGCCCTACCTGCTGATGGCT  372
                ||  ::::|:||:::    |||||||||||||           :::|||||||||::
Db     41  SerIleIleValThrValArgProGlySerAspProAspAlaGluProLeuThrLeuLeuGly    60

Qy    373  CACTTTGATCGTGGTTGCCCTCGCCCCTGAAGAGAAGGTGCCCCCATTCTTCTGGGTTG  432
                ||||||||||||||||                         ||||||||
Db     61  HisThrAspValArgProValAspSerLeuProLysTrpThrLysAspAspPheGlyAlaGlu  80

Qy    433  GAGCCGTGATGGCGTGTCATCTATGGTGCGGGGCGGCCACACTGGACGACAAGAACTCTGTGATGGCA  492
                |||||||  ||| :::|||||||||||::|||||  ||::||
Db     81  IleSerAspGlyGlnIleIleTrpPGlyArgGlySerValAsp-----------MetLeu  95

Qy    493  TTACTGCAGGCGCCTTGGGAGCTCTCCTCGTCGATCAGGAAGTACATCCCCGGAGATCT----  546
                :::|||                    :::|||::|::            |||
Db     96  PheIleThrAlaThrGlnAlaLeuAlaValThrArgGln----ValAlaArgGluGlyGlyLeu  114

Qy    547  --------------TTCTTCAATTCTCTGGGCCATGAATGATGAGGAGTCATCAGGGACACAGGGGCT  594
                              |||  :::   :::|||            :::|||
Db    115  ArgGlyThrLeuThrPheValGlyValAlaAspGluGlnLeuAlaArgGlyGlyLeuGlyAla  134

Qy    595  CAGAGGATCTCA--------------------------------GCCCTGCTACAGTCA  621
                :::  |:|||                                 |||
Db    135  LysTrpLeuSerGluGluHisGlnAsnLeuPheSerTrpLysAsnCysLeuSerGluSer  154

Qy    622  AGGGGCGTCCAGGCTAGCCCTTCATTGTGGACGAGGGGGGGGTTCAYCTTGGATGATTTCATT  681
                |||                       |||            :::|||
Db    155  GlyGlySerHisLeuPro--------------ValHisAspGly                    164

Qy    682  CCTAACTTCAAGAAGCCCCATCCGCCCTTGATTGCAGTGTCTCAGAGAAGGGGTTCCATGAACCTC  741
                   ::::|||                |||||||||||::::
Db    165  --------SerAspAlaValIleAsnValGlyGluLysGlyAlaAlaGlnAsnArg         180

Qy    742  ATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCCTCCT----CCAAGGAGAGACAGC  798
                :: |||||  |||||            |||||  |||         ||::||||::
Db    181  ArgIleHisValAsnGlyAspAlaGlyHisGlySerIleProPheAsnProArgAspSerAla  200

Qy    799  ATTGGCGATCCTTGCAGCTGCTTGTCAGCCCGATTGGAGCAGAGACACCAATGCCTATCATA-----  855
                ||| |||            ::|||                :::
Db    201  IleValLysIleGlyLysValAlaIleLeuArgArgIleAlaIleAlaAlaAsnAsnLysValAlaLys  220

Qy    856  ------TTTGGAAGCGGGACAGTGGTGGACYGTATTGCAGCAACTGGCAAATGAGTTT---  906
                      ||||||           |||                    |||
Db    221  AspAspIleArgProGlnGlyPheVal-----------------------GlnAlaHisArgPheAsp  234
```

```
Qy    907 CCCTTCCCTGTCAATATAAATCCTGAGC-------AACCCATGGCTATTTGAACCACTT---  957
              |||                          ::::|||||||                ::::|||
Db    235 ProGluThrGlnAlaLeuLeuLeuSerGlyThrSerProGluAlaThrAlaGluPheGly  254

Qy    958 ---ATAAGCAGGTTTANTGGAGAGAGAAAAATCCCTTAACCAAATGCAAATAATCAGGACCACCACG 1014
             :::|||||||||                      ::|||||||
Db    235 AlaHisAlaValSerHisLeuThrIle 268

Qy   1015 GCACTCACCATATTCAAAACGAGGGGGTCAAGTTCAAGTTCATCCCCCAGTGCCCCAGGCC 1074
              ||| |||||                              |||
Db    269 AlaGlnThrAlaValArgAlaGlyGlnAlaIleAsnValLeuProSerHisAlaThrYrLeu 288

Qy   1075 ACAGTCAACTTCCGGATTCACCCTGGACAGACA-------GTCCAAGAGGTCCTAGAA 1125
              ::::|||   |||                             |||
Db    289 GluLeuAspIleArgThrLeuProGlyGlnThrAsnAspThrYrValAspAspThrLeuArg 308

Qy   1126 CTCACGGAAGAACATTGTGGCTGATAACAGACTCCAGTTCCATNGTGTTGAGTGCCTTTGAC 1185
              :::|||||||                            :::
Db    309 AlaAlaLeuGlyAspLeuAlaAspGluValGlnIleGluHisLeuIleSer-----Glu 326

Qy   1186 CCCCTCCCCGTTCAGCCCTTGCATGACAACAAGGCCCTTGGGCGTTCACCAGCTGCTCCGGCCAGACC 1245
              ||||||||                              :::
Db    327 GluIleThrValSerProThrAspSerArgLeu-------TyrAsnThrLeuGlyLysVal 344

Qy   1246 GTACAGTCCGTCTTCCCGGAAGTCAATATTACTGCCCAGTTACTTCTATTGGAACACA 1305
              |||||||                              :::
Db    345 LeuGlyAspPhePheProAspAlaProVal---ValProIleIleSerSerGlyGlySer 363

Qy   1306 GACAGCCGATTCTTTACAAACCTCACCACTGGGCANTCTACAGGTTCTACCCCATCTACATA 1365
              |||||||                              :::
Db    364 AspLeuArgPheGlyLysArgLeu---GlyGlyValGlyTyrGlyPheAlaValHisAla 382

Qy   1366 CAG-------CCTGAAGACTTCAAAACGCCATCCATGGAGTGCAACGAG-------- 1404
              :::                                   |||
Db    383 ArgGluArgThrLeuAlaGlnAlaMetGlyGlnLeuHisSerHisSerAspGluAlaLeuTyr 402

Qy   1405 -------AAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAAA 1440
                          |||||||:::||||||:::
Db    403 LeuGluAspLeuGluLeuThrValArgGlyTyrAspSerValValArg 418
```

RESULT 8
AAB79819
ID   AAB79819 standard; Protein; 422 AA.
XX
AC   AAB79819;
XX
DT   30-APR-2001 (first entry)
XX
DE   Corynebacterium glutamicum MP protein sequence SEQ ID NO:372.
XX
OS   Corynebacterium glutamicum.
XX
PN   WO20010843-A2.
XX
PD   04-JAN-2001.
XX
PF   23-JUN-2000; 2000WO-IB00923.

PR   25-JUN-1999; 99US-0141031.
PR   01-JUL-1999; 99DE-1030476.
PR   02-JUL-1999; 99US-0142101.
PR   08-JUL-1999; 99DE-1031415.
PR   08-JUL-1999; 99DE-1031418.
PR   08-JUL-1999; 99DE-1031419.
PR   08-JUL-1999; 99DE-1031420.
PR   08-JUL-1999; 99DE-1031424.
PR   08-JUL-1999; 99DE-1031428.
PR   08-JUL-1999; 99DE-1031434.
PR   08-JUL-1999; 99DE-1031435.
PR   08-JUL-1999; 99DE-1031443.
PR   08-JUL-1999; 99DE-1031453.
PR   08-JUL-1999; 99DE-1031457.
PR   08-JUL-1999; 99DE-1031465.
PR   08-JUL-1999; 99DE-1031478.
PR   08-JUL-1999; 99DE-1031510.
PR   08-JUL-1999; 99DE-1031541.
PR   08-JUL-1999; 99DE-1031573.
PR   08-JUL-1999; 99DE-1031592.
PR   08-JUL-1999; 99DE-1031632.
PR   08-JUL-1999; 99DE-1031634.
PR   08-JUL-1999; 99DE-1031636.
PR   09-JUL-1999; 99DE-1032125.
PR   09-JUL-1999; 99DE-1032126.
PR   09-JUL-1999; 99DE-1032130.
PR   09-JUL-1999; 99DE-1032186.
PR   09-JUL-1999; 99DE-1032206.
PR   09-JUL-1999; 99DE-1032227.
PR   09-JUL-1999; 99DE-1032228.
PR   09-JUL-1999; 99DE-1032229.
PR   09-JUL-1999; 99DE-1032230.
PR   09-JUL-1999; 99DE-1032922.
PR   09-JUL-1999; 99DE-1032926.
PR   14-JUL-1999; 99DE-1032928.
PR   14-JUL-1999; 99DE-1033004.
PR   14-JUL-1999; 99DE-1033005.
PR   14-JUL-1999; 99DE-1033006.
PR   12-AUG-1999; 99US-0148613.
PR   27-AUG-1999; 99DE-1040764.
PR   27-AUG-1999; 99DE-1040765.
PR   27-AUG-1999; 99DE-1040766.
PR   27-AUG-1999; 99DE-1040832.
PR   31-AUG-1999; 99DE-1041378.
PR   31-AUG-1999; 99DE-1041379.
PR   31-AUG-1999; 99DE-1041380.
PR   31-AUG-1999; 99DE-1041394.
PR   31-AUG-1999; 99DE-1041396.
PR   03-SEP-1999; 99DE-1042076.
PR   03-SEP-1999; 99DE-1042077.
PR   03-SEP-1999; 99DE-1042079.
PR   03-SEP-1999; 99DE-1042086.
PR   03-SEP-1999; 99DE-1042087.
PR   03-SEP-1999; 99DE-1042088.
PR   03-SEP-1999; 99DE-1042095.
PR   03-SEP-1999; 99DE-1042124.
PR   03-SEP-1999; 99DE-1042129.
PR   09-MAR-2000; 2000US-0187970.
XX
PA   (BADI) BASF AG.
XX
PI   Pompejus M, Kroeger B, Schroeder H, Zelder O, Haberhauer G;
XX
DR   WPI; 2001-137957/14.
DR   N-PSDB; AAF71938.
XX
PS   Claim 20; Page 717-718; 1737pp; English.
XX
CC   AAF71753 to AAF72330 encode the Corynebacterium glutamicum metabolic
CC   pathway (MP) proteins given in AAB79634 to AAB80211. The C. glutamicum
CC   MP nucleic acids are useful for the production of fine chemicals
CC   in microorganisms, including organic acids, nonproteinogenic amino
CC   acids, purine and pyrimidine bases, nucleosides, nucleotides, lipids,
CC   saturated and unsaturated fatty acids, diols, carbohydrates, aromatic
CC   compounds, vitamins, cofactors, polyketides and enzymes.
XX
SQ   Sequence   422 AA;

PT   Nucleic acids from Corynebacterium glutamicum encoding metabolic
PT   pathway proteins, useful for producing fine chemicals in
PT   microorganisms, including organic acids, nonproteinogenic amino acids,
PT   and purine and pyrimidine bases -

```
Alignment Scores:
Pred. No.:            2.16e-17
Score:                274.00
Percent Similarity:   42.55%
Best Local Similarity: 25.48%
Query Match:          9.92%
DB:                   22

Length:       422
Matches:      106
Conservative: 71
Mismatches:   161
Indels:       78
Gaps:         20

US-10-014-896-1 (1-1509) x AAB79819 (1-422)

Qy  313 AGCCACCTGTTCACTATCAAAGGCTCGGACCCCAGCTTGCAGCCCTTACCTGCTGATGGCT 372
            |||    ||||||| :: ::: ||
Db   41 SerIleIleValThrValProGlySerAspAspAlaGluProLeuThrLeuGly 60

Qy  373 CACTTTGATGTGGCTTCGCCTGCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTCTGGGTTG 432
            |||   :::||
Db   61 HisThrAspValValProValAlaLeuProLysTrpThrLysAspProPheGlyAlaGlu 80

Qy  433 GAGCGTGATGCGTCACTCTATGGTCGGGGACTCCGACCACTGGACGACAAGAACTCTGTGATGGCA 492
            ||||| :::|||::|||||:::||||:::||||
Db   81 IleSerAspGlyGlnIleThrProGlyArgGlySerValAsp 95

Qy  493 TTACTGCAGGCCCTTGGAGCTCCTGCTGCTGATCACTACACATCCCCCGAGATCT------ 546
            ::||| ::::|| |||:::
Db   96 PheIleThrHisArgLeuAlaAlaValThrArgGln---ValAlaArgGluGlyLeuLeu 114

Qy  547 ------TTCTTCATTTCTTCTGGGCCCATGATGAGGAGTCATCAGGACAGGGGCT 594
            ::|||   :: |||
Db  115 ArgGlyThrLeuThrPheValGlyValAlaLaAspGluValGlyLeuGlyAla 134

Qy  595 CAGAGGAATCTCA---------GCCCTGCTACAGTCA 621
            ::: ||||       :::||   ::|||
Db  135 LysTrpLeuSerGluGluHisGlnAsnLeuPheSerTrpLysAsnCysLeuSerGluSer 154

Qy  622 AGGGGGGTCCAGCTACCCTTCATTGTGGGACGAGGGGGGTTCATCGTTGGATGAATTCATT 681
            ||   :::|||
Db  155 GlyGlySerHisLeuPro------ValHisAspGly 164

Qy  682 CCTAACTTCAAGAAGCCCATCGCCTTGATCTGGCAGTCTCAGAGAGGGTTCCATGAACCTC 741
            ||||||||||||||:::
Db  165 ------SerAspAlaValValIleAsnValGlyGluLysGlyValAlaLaGlnArg 180

Qy  742 ARGCTGCAAGTAAACATGACTTCAGGCCCACTCTTCAGCTCCT---CCAAGGAGACAAGC 798
            ||||||   :::|||  ::|||
Db  181 ArgIleHisValAlaAsnGlyAspAlaGlyValHisGlySerIleProPheAspSerAla 200

Qy  799 ANTTGGGATCCATCCCTGCAGCTGCGTGTCACGGGATTGGGACGACGCAGCACCCATATCATA--- 855
            ::: :::|||||||||||||
Db  201 IleValLysSerIleGlyLeuValAlaLaArgGlyIleAlaLaAsnLeuAspAspLeuLys 220

Qy  856 ------TTTGGAAGCGGGACAGTGGTCACTGTATTGCAGCAACTGGCAAATGAGTTT--- 906
            |||   :::|||  ||
Db  221 AspAspIleTrpGlnGlyPheVal------GlnLeuHisArgPheAsp 234

Qy  907 CCCTTCCCTGTCAATATAATCCTGAGC-----AACCCATGGCTATTTGAACCACTT--- 957
            :::||||| :: |||
Db  235 ProGluThrGlnLnAlaLeuLeuSerGlyThrSerProGluAlaTyrAlaGluProGly 254

Qy  958 ---ATAAGCAGGTTTATGGAGAGAAAATCCCTTAACATATCAGATAACTCAGACCACACG 1014
            :::||||||||:::
Db  255 GlyLeuSerArgPhe------AlaHisAlaValSerHisIleLeuThrIle 268

Qy 1015 GCAGCTCACCATATTCAAAGCAGGGTCAAGTTCAATGTCATCCCCCCAAGTGGCCCAGGCC 1074
            |||   :::||| :::||
Db  269 AlaGlnThrValAlaArgAlaGlyGlnAlaLaIleLeuAsnValLeuProSerHisAlaTyrLeu 288

Qy 1075 ACAGTCAACTTCCGGATTCACCCTGCCGACAGACA-----GTCCAAGAGGTTCCTAGAA 1125
            |||   :::|||   |||
Db  289 GluLeuAspProLeuArgThrLeuProGlyGlnThrAsnAspProTyrValAspAspThrLeuArg 308

Qy 1126 CTCACGAAGAAACATTGTGGCTGATAAACAGAGTCCAGTTGCCAATCCAGTGTGCCTTTGAC 1185
            ::|||:::||||
```

```
Db  309 AlaAlaLeuGlyAlaAspLeuAlaAspGluValGluValGluHisLeuIleSer------Glu 326
            ||| ||||||| :: ::: ||
Qy 1186 CCCCTCCCCGTCAGCCCCTTCGATGACAAGCAAGGCCTTGGGGTTACCAGCCTGCCCAGACC 1245

Db  327 GluAlaThrValSerProThrAspSerArgLeu------TyrAsnThrLeuGluLysVal 344
            ::: :::||| :::
Qy 1246 GTACACGTCCGTCTTCCCGGAAGTCAATATTACTGCCCAGTTACTTCTATTGGAACACA 1305

Db  345 LeuGlyAspPhePheProAspAlaProVal---ValProIleIleSerSerGlyValSer 363
            :::||| :::
Qy 1306 GACAGCCGATTCTTTACAAACCTCACCACTGGCATCTACACAGGTTCTACCCCATCTACACA 1365

Db  364 AspLeuArgPheGlyArgArgLeu---GlyGlyValGlyTyrGlyPheAlaValHisAla 382
            ||| :::||| :::
Qy 1366 CAG------CCTGAAGACTTCAAACGCACCATCCATGGAGTCCAACCAG------ 1404

Db  383 ArgGluArgThrLeuAlaMetGlyGlnLeuHisSerHisAspGluAlaLeuTyr 402
            ||| :::||| :::
Qy 1405 ------AAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAA 1440

Db  403 LeuGluAspLeuGluLeuThrValArgGlyTyrArgAspSerValValArg 418
            :::||| :: ::|||
```

RESULT 9

AAU41311

ID   AAU41311 standard; protein; 458 AA.
XX
AC   AAU41311;
XX
DT   13-FEB-2002 (first entry)
XX
DE   Propionibacterium acnes immunogenic protein #2207.
XX
KW   SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW   uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW   inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW   dermatological; osteopathic; neuroprotectant.
OS   Propionibacterium acnes.
XX
PN   WO200181581-A2.
XX
PD   01-NOV-2001.
XX
PF   20-APR-2001; 2001WO-US12865.
XX
PR   21-APR-2000; 2000US-1990047P.
PR   02-JUN-2000; 2000US-208841P.
PR   07-JUL-2000; 2000US-216747P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
PI   L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX
DR   WPI; 2001-616774/71.
DR   N-PSDB; AAS59514.
XX
PT   Propionibacterium acnes polypeptides and nucleic acids useful for
PT   vaccinating against and diagnosing infections, especially useful for
PT   treating acne vulgaris -
XX
PS   Example 1; SEQ ID NO 2506; 1069pp; English.
XX
CC   Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC   polypeptides. The proteins and their associated DNA sequences are used in
CC   the treatment, prevention and diagnosis of medical conditions caused by
CC   P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC   pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
CC   P. acnes is also involved in infections of bone, joints and the central
CC   nervous system, however it is particularly involved in the inflammatory
CC   lesions associated with acne vulgaris. A method for detecting the
CC   presence or absence of P. acnes in a patient comprises contacting a
CC   sample with a binding agent that binds to the proteins of the invention

CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA).
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence   458 AA;

Alignment Scores:
Pred. No.:   1.2e-16
Score: 266.50
Percent Similarity: 41.32%
Best Local Similarity: 25.71%
Query Match: 9.65%

	Length:	458
Matches:	117	
Conservative:	71	
Mismatches:	212	
Indels:	55	
Gaps:	18	

US-10-014-896-1 (1-1509) x AAU41311 (1-458)

Qy   175 ATCCAGAATTCCAAACACTGACTTTTAGCTCTCTGAGAAGTCCAAT---------------- 216
           |||||||||||| ::::   :::
Db   35 IleGlnIleAspSerGlnAsnPheGlyProGlnAspAlaArgGlyGluValGlumetCys   54

Qy   217 -----ACTACAGCCCTGGGCTGAG--------------TTCGGAAAAATACATTCATAAAGTCTTT 261
                ::: |||||| :::||:|||||
Db   55 HisTyrValThrGlyLeuLeuAspGluValGlyValThrLeuHisGluSerGlu 74

Qy   262 CCTACAGTGGTCGCAGCACCAGCTTATCCAGCAAGTCGTGGAAGAGTATAGCCACCTG 321
           :::||:| ||| ||:||:|||||:||
Db   75 ProGlyArgValThr----------------LeuValAlaGluThrpAlaPro--- 86

Qy   322 TTCACTATTCCAAGGGTCGGACCCCAGCTTGCAGGCCTACCTGCTGATGGCTCCACTTGAT 381
                :::|:||:||:|| ||||| |||| |||||||| ||||
Db   87 ------------GluGlyThrAspThrSerArgProAlaLeuLeuLeuHisGlyHisSerAsp 103

Qy   382 GTGGTGCCTGCCCCTGCCAAAGAAGCTGCGGAGGTGCCCCAATTCTCTGGGTTGGAGGGTGAT 441
           |||||||| |||||||||| ||||| ||||||||||||||||||||
Db   104 ThrValProPheGluAlaLeuAlaAlaSPrpThrHisHisProLeuSerGlyIleHisAsp 123

Qy   442 GGGCGTCATCTATATGGTCGGGGGGCACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAG 501
           ::: ||||||
Db   124 AsnCysValTrpGlyValArgGlyAlaIleAspMetLysGlyPheLeuLeuAlaMetValLeuSer 143

Qy   502 GCCTTGGAGCCCTGCCTGATCAGGAAGAAGTACATCCCCGAGAGATCTTTCTTCATTTCTCTG 561
                            ||| :::||||
Db   144 AlaIleArgAlaArgGlnArgGlyArgGlyGluValProSerArgProIleArgPheIleMet 163

Qy   562 GGCCATGATGAGGAGAGTGTCATCAGGGACA---------------GGGGCTCAGAGGATC 603
           ||||||||||| ||||||||
Db   164 PheAlaAspGluGluCysSerSerGlyThrLeuGlyAlaThrHisPro 183

Qy   604 TCAGCCTGCTACAGTCAAGGGGGTGCCAGCTAGCCTTCATTGTGGACGAGGGGGCTTC 663
           ||||||
Db   184 GluAlaPhe--------AspGlyValThrGluAla---IleSerGluValGlyValGlyPhe 199

Qy   664 ATCTTGGATGATGATTTCATTCCTAACTTCAAGAGACCCATCGCCTTGATTGCAGTGTCAGAG 723
           ::: ||| :::||:
Db   200 SerLeuThr-----ThrProGlnGlyLysArg---ValTyrValIleGlnIleSerAlaGlu 216

Qy   724 AAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGCCCACTCTTCAGCTCCT 783
           ||||||| :::
Db   217 LysGlyLeuTrpIleProheargMetSerAlaThrGlyValHisGlySerMetArg 236

Qy   784 CCAAAGGAGACAACAGCAATTGGCATCCTTGCAGCTGCTGTCACACCGATTGGAGCAGACACCA 843
           ::|
Db   237 AsnProAspAsnAlaValThrArgValLeuAspAlaLeuSerArgIleAspSerTyrGln 256

Qy   844 ATGGCCTATCATATTTTGGAAGCGGGACAGTGGTCGACTGTATTGCAGCAACTGGAAAATGAG 903
           ||| :::   ::|

Db   257 TrpProAspLeuHisHis------------------ProValGlnGluGluPheLeuAsnGln 271

Qy   904 TTTCCCTTCCCGTGTCAATATATAATACCTGAGCAACCCAACCCCATTGGCACACTTATAAGC 963
           ||| ::: :::   ::::

Db   272 ValAlaAlaMetTrpGlyTyrLeuThrIleLeuAspPheArgAspAspLeu---GluSerSerLeuSer 290

Qy   964 AGGTTTATGGAGGAGAAATCCCTTAACCAATGCAATAATCAGGACGAACCACCACCACTCACC 1023
           :::||| |||

Db   291 ProIleGlySerLeuSerArgMetValAlaAlaLeuAlaCysSerAlaHisAsnValThrProThr 310

Qy   1024 ATATTCAAAGCAGGGGGTCAAGTTCAATGTAATGTCATCCCCCAGTGCCCAGGCCCACAGTCAAC 1083
           :::|||||| ||||||

Db   311 ValLeuSerAlaGlyTyrLysValAsnValValProThrArgAlaSerArgAlaGluValAsp 330

Qy   1084 TTCCGGATTCACCCTGGACAGCAGACAGGTTCCTAGAACTCACGAAGAACATTGTG 1143
           :::||| :::||:

Db   331 AlaAlaArgPheIleProGly----AlaGluGluAspAspMetIleLeuGlyHisLysSerLeu 348

Qy   1144 GCTGATAACACAGATCTTCCATGTTGTGTTGAGGTGCCTTTGACCCCCTCCCCAGGCTCAGCCCT 1203
                               |||

Db   349 AlaGlyProGlyIleGluAspPheGluThrIleSerArg-----LysProAlaThrAlaAlaPro 367

Qy   1204 TCTGATGACAAGGGCCCTTGGGGTCTACCAGCTGCTCCGGCCGACAGAGTCCGTGTTCTTCCCG 1263
           :::   :::

Db   368 PheGluGlyAlaAlaVal-----AspAlaIleArgArgAlaValAlaAspAlaValAspPro 385

Qy   1264 GAAGTCAATATTACTGCCCCCAGTTACTTCTAATTGGCAACACAGACAGCCAGATTCTTTACA 1323
           :::   ::|

Db   386 GlyAla---ValValLeuProTyrLeuAsnSerAlaGlyThrArgAspAlaGlyPheAla 404

Qy   1324 AACCTCACCACTGGC----------------ATCTACAGGTTCTCACCCTACTACATACAGGCCT 1371
           |||

Db   405 ValLeuProAspGlyArgArgIleAsnTyrTyrGlyCysThrProLeuArgLeu---Pro 423

Qy   1372 GAAGACTTCAAAACGCATC---------------CATGGGAGTCAACGAAGAAAATCTCAGTCCAAGCC 1422
           :::||:||| :::||:

Db   424 AlaAspPheAspPheIleAsnLeuPheHisGlyValAspArgValProValGlySer 443

Qy   1423 TATGAGACCCAAGTGAAAATTCATCTTTGAGGTTGATTCAGAAATGCT 1467
           |||

Db   444 LeuValPheGlyAlaLysValAlaValAspHisIleLeuGlnGluAla 458

RESULT 10
ABB62639
ID   ABB62639 standard; Protein; 401 AA.
XX
AC   ABB62639;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 14709.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li FWD, Myers EW;
XX
DR   WPI; 2001-658860/75.
DR   N-PSDB; ABI06742.
XX

PT New Isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 14709; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signaling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL161675), expressed DNA
CC sequences (ABL16176-ABL30511), and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 401 AA;

Alignment Scores:
Pred. No.:   2.02e-15
Score: 253.50
Percent Similarity: 38.77%
Best Local Similarity: 24.58%
Query Match: 9.18%
DB: 22

Length: 401
Matches: 116
Conservative: 67
Mismatches: 180
Indels: 109
Gaps: 19

US-10-014-896-1 (1-1509) x ABB62639 (1-401)

```
Qy  130 AGCAAAGAGGAACGGCTCGGCGATGAAAGAGGCGTGAAAGTGCCATCCAGAATTCCAACA 189
            |||||| |||||| |||||||  ||
Db    8 AsnAsnGluGlnIleLysIleProHisArgGluThrLeuArg------IleProThr 23

Qy  190 GTGACTTTTAGCTCTGAGAAGTCCAATACTACACAGCCCTGGCTGGGAGAAAATAC--- 246
            ::: ||||||  :::||  |||||
Db   24 Val------HisProAspValAspThrTyrThrAlaCysValGluPheLeuLysArgGln 40

Qy  247 ----------ATTCATAAAGTCTTCACCTACAGTGGTCAGACCACCAGCTTT 285
                       ||||||
Db   41 AlaSerSerLeuAsnLeuProValGlyLeuValValTyrProAlaAlaValGlnThrLysProVal 60

Qy  286 ATCCACGACATGAAGTCCGTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCC 345
            |||
Db   61 Val------IleIleLysTrpGluGlySerGlnPro 70

Qy  346 AGCTTGCAGCCCTACCTGCGATGGCCTCACTTTGAGGCGGTCGCCCTGCGCCCGAAGAAGGC 405
            ::|||  |||||||||||
Db   71 GluLeuSerSerIleValLeuAsnSerHisThrAspValValProValPheArgGluLys 90

Qy  406 TGGGAGGTGCCCCCATTCTCTGGG--------TTGGAGCGTGATGGCGTCAATCTATGGTCGGGGC 462
            :::
Db   91 TrpThrHisGluProPheSerAlaAspIleGluGlyArgIlePheAlaArgGly 110

Qy  463 ACACTGGACGACAAGAACTCTGATGATGGCAATTACTCGCAAGGCCTTGGAGCTCCTGCTGATC 522
            |||   |||   :::   |||:::   ||||||
Db  111 ThrGlnAspMetLysSerValGlyThrGlnTyrLeuGlnAlaIleArgGlnLeuLeuLysAla 130

Qy  523 AGGAAGTACATCCCCCAAGAGATCTTTCTTCATTCTCTGGGCCATGAGTGAGGAGTCATCA 582
            |||:::|||:::|||   ||||||||
Db  131 SerGlyPheLysProLysArgAsnLeuThrValThrPheGluValProAspGluThrGly 150

Qy  583 GGG---ACAGGGGCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGCGTCAGCTAGCC 639
            |||   ||||||:::||||||||
Db  151 GlyHisLeuGlyMetAlaGluPheValLysThrAspTyrThrTyrLysMetAsnAlaGly 170

Qy  640 TTCATTGTGACGAGGGGGGGCTTCACTCTGGATGATTCATTCCTAACTCAAGAAGCCC 699
            |||   |||||||||||||
Db  171 PheSerLeuAspGluGlyAlaThrSerGluSerAsp 182

Qy  700 ATCGCCTTGATTGCAGTCTCAGAAGAGGGTTCCATGAACCTCATGCTGCAAGTAAAACATG 759
            :::   ||| ||||||||||||
Db  183 ValHisHisSerLeuPheTyrAlaGluArgLeuArgTrpGlyLeuLeuLysLeuLysValSerGly 202

Qy  760 ACTTCAGGCGACTCTTCAGCTCCTCCAAAGGAGGACAAGGATTGGCATCCTTGCAGCTGCT 819
            ||||||||||||||  :::   |||       |||
Db  203 ThrSerGlyHisGlySerLeuLeuLeuProAsnThrAlaGlyValLeuLysLeuAsnTyrLeu 222

Qy  820 GTCAGCCGATTGGAGCAGACCAACGCCAATGCCTATCATCATATTTGGAAGCGGGGACAGTGGACT 879
            :::   |||:::   |||          :::   |||
Db  223 ValAsnLysLeuThrGlu------PheArgThrSerGlnVal- 234

Qy  880 GTATTGCAGCAAACTGGCAAATGAGTTTGCCCTTTCCCTTGTCAATATAATCCTGAGAACCCA 939
            :::|||  |||:::
Db  235 ------GluAsnLeuAlaArgAsp- 240

Qy  940 TGGGCTATTTGAACCACTTAATAAGCAGGTTTTATGGGAGAGAAATCCCTTTAACCAATTGCAATA 999
            :::
Db  241 --------SerSerLeuSerLysGlyAsp 247

Qy 1000 ATCAGGAGACCACCACGGCCACTCCACCATATTCAAAAGCAGGGGGTCCAAGTTCAATGTCATCCCC 1059
            |||:::   |||   |||:::       |||
Db  248 Val--ThrThrValAsnLeuThrGlnLeuSerGlyLysGlyValGlnSerAsnValValPro 266

Qy 1060 CCAGTGGCCCAGGCCACCAACTTCCGGATTCACCCTGGACAGCAGACAGTCCAAGAGGTC 1119
            |||:::|||:::   |||
Db  267 ProLeuProHeGluValProGlySerTyr------Phe 277

Qy 1120 CTAGAACTCACGAAGAACATGTGGGTGATAACAGA------ 1155
            ::|||   |||||||||||
Db  278 --AlaIleThrValAsnValAlaIleAlaPheGluLysGlnIleIleArgAspTrpCysGluGlu 296

Qy 1156 ------------GTCCAGTTCCATGTGTTGGAGTGGTCCTTGACCCTCCCC--- 1194
            |||
Db  297 -----------PheGlnLysGluProTyr 311

Qy 1195 GTTCAGCCCTTCTGAT--GACAAGGCCTTGGGCTACCAGTGCTCCGGCGCCGTACAG 1251
            ||||||:::|||       :::   |||
Db  312 IleGlyProThrLysGluAsnSerAsnProTyrTrpLeu------AlaValLys 328

Qy 1252 TCCGTCTTCCCCGGAGTCAATATTACTGCC---CCAGTTACTTCTATTGGCAAACACAGAC 1308
            |||||||||||  :::|||||||
Db  329 AlaAlaIleLeuAspGluLeuGlyLeuGlyValHisProIleValCysProGlyAlaThrHrsAsp 348

Qy 1309 AGCCGGATTCTTTACAAACCTCACCACGGGCATCTACAGGTTCTAACCACTACACAACAG 1368
            |||:::   |||   |||:::
Db  349 SerArgPheIleLeuArgGluLysGlyThrProAlaIleGlyPheSerProIleIle--- 366

Qy 1369 CCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAAAATCTCAGTCAGTCCCTATGAG 1428
            |||:::   |||   |||:::
Db  367 ---AsnThrThrMetArgIleHisAspHisIleAspGluPheLeuGlnAlaAspValTyrLeu 385

Qy 1429 ACCCAAGTGAAATTCATCCTTTGAGTTGATTCAGAAT 1464
            :::   |||:::|||:::|||
Db  386 AsnGlyIleAspValTyrLysIleIleIleArgAsn 397
```

RESULT 11
ABB66170
ID   ABB66170 standard; Protein; 400 AA.
XX
AC   ABB66170;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 25302.
XX
KW   Drosophila; developmental biology; cell signaling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.

XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL10273.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Disclosure; SEQ ID NO 25302; 21pp + Sequence listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC sequences (ABL16176-ABL30511), and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 400 AA;

Alignment Scores:
Pred. No.:
Score: 228.00
Percent Similarity: 38.89%
Best Local Similarity: 23.99%
Query Match: 22

DB: Length: 400
                        Matches: 95
                        Conservative: 59
                        Mismatches: 158
                        Indels: 84
                        Gaps: 13

US-10-014-896-1 (1-1509) x ABB66170 (1-400)

Qy 334 GGCTCGGACCCCAGCTTGCAGCCCTTACCTCCTGATGGCGTCCACTTTGATGGTGCCTGCC 393
            |||::||||:||:||| ||| :: |||| |||||||||||||
Db   66 GlyThrGluProGluLeuProSerValIleLeuAsnSerHisMetAspValValProVal 85

Qy 394 CCTGAAGAAGGCTGGGAGGTGCCCCCATTCTCTGGG---TTGGACGGTGGATGGGGTCATC 450
            |||| ||| :: ||||
Db   86 PheArgAspIleTrpThrHisAsnProPheAlaAlaAspIleAspGluGluGlyArgIle 105

Qy 451 TATGGTCGGGGCACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAGGCCTTGGAG 510
Db 106 PheAlaArgGlyThrGlnAspMetLysSerValGlyThrGlyThrTyrLeuGlyAlaIleArg 125

Qy 511 CTCCTGCTGATCAGGAAGTACATCCCCCAAGAATCTTTCTTCATTTCTCTGGGCCATGAT 570
Db 126 LeuLeuLysAlaSerGlyValGlnProLysArgAsnPheGluThrPheValThrPheAsp 145

Qy 571 GAGGAGTCATCAGGG---ACAGGGGCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGC 627
Db 146 GluGluIleGlyGlyLeuGlyMetGlnGluProPheValLysThrGlyTyrTyrSerAsn 165

Qy 628 GTCCAGCTAGCCTTCATTGTGGACGAGGGGCTTCACTGGATGAATGTTCATTCCTAAC 687
Db 166 MetAsnValGlyPheSerLeuAspGluGlyLeuGlyThrSerGluIleAsp------- 181

Qy 688 TTCAAGAAGCCCATGCCTTGAATTGACGAGAGGGTTCCATGAACCTCATGCTG 747
Db 182 --------LeuPheTyrValPheTyrAlaGluArgMetArgTrpGlyLeuLysLeu 197

Qy 748 CAAGTAAACATGACTTCAGGCCACTCTTCAGCTCCTCCAAAGAGACAAAGCAATGGCATC 807
Db 198 AsnPheSerGlyThrSerGlyHisSerGlySerMetLeuLeuProSerThrAlaGlyValLys 217

Qy 808 CTTGCAGCTGCTGTGTCAGCCGATTGGAGCAGAGACACCAAATGCCTATCATATTGGAAGCGGG 867
            ||| :||
Db 218 LeuAsnThrValLeuAsnLysLeuThrGlu---------------------PheArgGluSer 231

Qy 868 ACAGTGGTGACTGTGATTGCAGCAACTGGCAAATGAGTTTCCCTTGCCTGTTCAATATAATC 927
            :||
Db 232 GlnVal---------------GlnArgLeuAlaAsnAspGlnThrIleAsnIleGlyAspVal 247

Qy 928 CTGAGCAACCCATGGCTATTTGAACCACACTTTATAAGCAGGTTTATAGGAGAGAAAATCCCTTA 987
            :::
Db 247 --- 247

Qy 988 ACCAATGCAATAAATCAGGACCACACCACGGGACTCACCATATTCAAAGCAGGGGTCAAGTTC 1047
            |||:||||::|
Db 248 --------------------ThrThrIleAsnLeuThrGlnLeuSerGlyValValGlnSer 261

Qy 1048 AATGTCATCCCCCAGTGGCCCAGCCACAGTCAACTTCCGGATTCACCCTGGACAGACA 1107
             |||||||:::||||||||
Db 262 AsnValValProProHisPheGluAlaValPheAspMetArg----------------- 275

Qy 1108 GTCCAAGAGGTCCTAGAACTCACGAAGAACATTGTTGGGCTGACATAACAGAGTCCAGTTCAT 1167
             || |||||||||||||||
Db 276 ------------LeuSerIleThrLeuAspValValAlaPheGluLys---GlnIleHis 290

Qy 1168 ------------------------------GTGTTGAGTGCCTTTGAC------ 1185
                                          :::||||||
Db 291 AspTrpCysGluGluValAlaGlyGlyGlyIleGluIleSerPheAspGlyLysProTyr 310

Qy 1186 ------CCCCTCCCGTCAGCCCTTCTGATGACAAAGGCCTTGGGCTACCAGTCCGGC 1239
            ||| :::
Db 311 ValAlaAlaProThrLysIleAspAspSerAsnProPheTrpLeuValAlaPhe---- 326

Qy 1240 CAGACCGTACAGTCCGTCTTCCCGGAAGTCAAATATTACTGCC---CCAGTTACTTCTATT 1296
            |||:::||||||
Db 327 ------------------LysAsnThrThrMetArgIleHisAspHisAspGlyPheLeuGlyAla 343

Qy 1297 GGCAACACAGACAGCCGAGTTCTTTACAAACCTACCACTGGCAATCTACAGGTTGTCTACCC 1356
            ||||||||||||||||
Db 344 GlyAlaThrAspSerArgAsnIleArgAlaLeuGlnGlyTyrIleProAlaLeuGlyPheSerPro 363

Qy 1357 ATCTACATTACAGCCTGAAGACTTCAAACGCCATCCATGGAGTCAAGGAAAAAACTCAGTC 1416
            |||||||
Db 364 Ile------------------LysAsnThrThrMetArgIleHisAspHisAspGlyPheLeuGlyAla 380

Qy 1417 CAAGCCTAAGAGACCAACAGCCGAGTTCTTTACAAACCTACCACTGGCAATCTACAGGTTGTCTACCC 1464
            ||||:::
Db 381 GluThrThrTyrLeuLysGlyIleGlnIleIleGlyTyrLysLysIleLeuGlyAsn 396

RESULT 12
ABB66172

ID ABB66172 standard; Protein; 408 AA.
XX
AC ABB66172;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 25308.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.

```
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL10275.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 25308; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signaling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175), expressed DNA
CC   sequences (ABL16176-ABL30511), and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   408 AA;

US-10-014-896-1 (1-1509) x ABB66172 (1-408)

Alignment Scores:
Pred. No.:            1.3e-12
Score:               224.50         Length:        408
Percent Similarity:   36.27%        Matches:       108
Percent Identity:     23.18%        Conservative:   61
Best Local Similarity: 8.13%        Mismatches:    182
Query Match:          22            Indels:        115
DB:                                 Gaps:           17
```

```
Qy      136 GAGGAACGCGTCGCGATGAAAGAGAGGCGGCTGAAAGTGCATCCAGATTCCAACAGTGACT 195
                |||||  :::|||  |:||:||:|||||  ||||:|| :|||||||||||||:
Db       12 GAAGTGATCATCCTGAAGATGTGCCCCAATTCTCTGGG---TTGGAG 435

Qy      196 TTTAGCTCTGAGAAG---------------------------TCCAATAACTACA 222
                ::|||  ||| |:|                           ||| :||
Db       32 TyrSerLysAspLeuPhePheArgSerLeuAspPheValIleSerLeuAlaAla 51
```

```
RESULT 13
ID   ABB62643
AC   ABB62643;
DT   26-MAR-2002   (first entry)
DE   Drosophila melanogaster polypeptide SEQ ID NO 14721.
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
OS   Drosophila melanogaster.
```

XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US09231.
XX
PR   23-MAR-2000; 2000US-191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI: 2001-656860/75.
DR   N-PSDB; ABL06746.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions -
XX
PS   Disclosure; SEQ ID NO 14721; 21pp + Sequence listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   340 AA;

Alignment Scores:
Pred. No.:
Score: 6.25e-12
Percent Similarity: 217.00
Best Local Similarity: 41.01%
Query Match: 23.88%
DB: 7.86%
                       22

US-10-014-896-1 (1-1509) x ABB62643 (1-340)

Db   109 ArgGlyAlaGlyLnAspMetLysSerValGlyThrGlnTyrLeuGlyAlaIleArgLeuLeu 128

Qy   124 CAGTTCAGCAAAGAGGAACGCGTCGCGATGAAAGAGGCGCTGAAAGGTGCCATCCAGATT 183
              ::::: :: ||| ||| :: |||| ::: ||| |||::||:||
Db   5 LysTrpGluLysAsnGluGluIleArgIle------PheArgGluTyrLeuArgIle 21

Qy   184 CCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGAAAA 243
              |||:|||:|| ||||| ||| :: ||| |||||| | : ||| || || ::|||
Db   22 ProSerVal------HisProAsnIleAspTyrThrAlaCysValGlnPheIleLys 38

Qy   244 TAC---------------------ATTCATAAAGTCTTTCCTACAGTGGTCAGCAC 279
                                      ::| |||||||||| | ||| |||||| |||
Db   39 ArgGlnAlaAspSerLeuAsnLeuProValGluValPheProAlaValLysSerLys 58

Qy   280 AGCTTTATCCAGCATGAAAGTCGTGGAAGAGTATAGCCACCTGTTCACTAACAGGCTCG 339
              :::   :||| : |||:::::: |||| ||
Db   59 ProValVal------------------IleIleLysTrpGluGlySer 68

Qy   340 GACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGTGCTGCCCCTGAA 399
              ||| |||||| | ||| ||| || || |||||||||||| || || || ||| |||
Db   69 GlnProLysLeuProSerIleIleLeuSerSerHisMetAspValValProValPhePro 88

Qy   400 GAAGGCTGGGAGGTGCCCCCATTCTCTGGG---TTGGACGTGAATGGCGTCATCTATGGT 456
              || |||||| ||||| || |||| ||| ||| | ||| ||| || ||||| |
Db   89 GluMetTrpThrHisGluProPheSerAlaAspIleAspGluGlyArgIlePheAla 108

Qy   457 CGGGGCACACTGGACGACGAAGAACTCTGTGATGGGCATTACGCCCTTGGAGCTCCTG 516
              ||||| | ||| || || |||| ::| || || | ||| ||| ||

```
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL06735.
XX
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PI   interactions -
XX
PS   Disclosure; SEQ ID NO 14688; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins
CC   (ABB57737-ABB72072).
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence   359 AA;
```

```
US-10-014-896-1 (1-1509) x ABB62632 (1-359)

Alignment Scores:
Pred. No.:
Score:                  5.34e-11
Score:                  207.50
Percent Similarity:     41.51%
Best Local Similarity:  25.79%
Query Match:            7.51%
DB:                     22

Length:        359
Matches:       82
Conservative:  50
Mismatches:    113
Indels:        73
Gaps:          12
```

```
Qy   214 AATACTACACAGCCCTGGCTGAGTTCGGAAAATACATTCATTAAAAGTCTTTCCTACAGTG--- 270
             |||          :::    |  ||||||        ::|:|||:::
Db     8 AsnAsnGluGlyIleLysIlePheArgGlyTyrLeuArg-----IleProThrValHis 25

Qy   271 ------------GTCAGCACCAGCTTTATCCAGCATGAAGTC------------ 300
                       |||||||||  |  |||||||||||
Db    26 ProAsnValAspTyrThrValCysThrSerThrGluPheLeuLysArgGlnAlaAlaSerLeuAsp 45

Qy   301 ----------------GTGGAAGAGTATAGCCACCGTGTTCACTATC--- 330
                             |||       ::|
Db    46 LeuProValGluValIleTyrProValAsnGluAsnProValValValIleLeuLysTrp 65

Qy   331 CAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTCGATGGCGTCACTTTGATGTGGCCT 390
             |||||||||  |:||      ::|||   |||||||||||
Db    66 GlnGlySerGlnProGluLeuProSerIleIleLeuAsnSerHisThrAspValPro 85

Qy   391 GCCCCTGAAGAAGGCTGGGAGGTGCCCCAATTCTCTGGG---TTGGAGCGTGAATGGCGTC 447
             |||||||||||||||||||||              :||||
Db    86 ValPheGluGlyTrpThrHisGlyProPheSerAlaAspLeuAlaGluGlyArg 105

Qy   448 ATCTATGGTCGGGGGCCACTGGACGACGAGAACTCGTGATGGAGCAATACTGCAGGCCTTG 507
             |||||||||  |||:||||||||                |:|||
Db   106 IlePheAlaArgGlySerGlnAspMetLysCysValGlyThrGlnTyrLeuGlyAlaVal 125

Qy   508 GAGCTCTGCTGATCAAGGAAGTACATCCCCGGAAGANTCTTTCTTCATTTCTCTGGGCCAT 567
             |||||||||  ||  ::|      ::|:||:||:||:
Db   126 ArgAlaLeuLysAlaSerGlyTyrGlnProLysArgThrIleTyrIleuThrTyrValPro 145

Qy   568 GATGAGGAGTCATCAGGGAGCAGGGGGCTCAGGGATCTCAGCCCCTGCTACAGTCA----- 621
             ||||||||||                  |||||||::|:
Db   146 AspGluGluValGlyLysHisLeuGluGlyMetArg-----GluLeuValLysSerAspTyr 163

Qy   622 ---AGGGGCGTCCAGCTAGCCTTCAATTGTGGACGAGGGGGGGCTTCATCTCTTGGATGATTTC 678
                ::::                      |||||||||
Db   164 PheLysLysLeuAsnValGlyPheSerPheAspGluGly----------------------- 176

Qy   679 AATTCCTAACTTCAAGAAGAGCCCCATCGCCTTGATTGCAGTCTCAGAGAAGGGTTCCATGAAC 738
             |||||||  :::           ||||||::|           |||||::
Db   177 IleSerSerAlaAspGluThrTyrAlaLeuTyr---TyrAlaGluArgThrLeuTrpHis 195
```

```
Qy   739 CTCATGGCTGCAAGTAAACATGACACTTCAGGCGCCACTCGTTCAGCTCCTCCAAAGGAGACAAGC 798
             |||                 ::::::|      |||::|||||        |||:::
Db   196 LeuArgPheLysIleSerGlyThrAlaGlyHisGlySerLeuLeuLeuProAsnThrAla 215

Qy   799 ATTGGCATCCTTGCAGCTGGTCGCTGCTTGTCAGCGCGCCGATTGGAGCAGCAACACCAAATGCCTATCATATT 858
             |||||||||||||||         :||::|||::          ||||::|::||:||:::
Db   216 GlyGluGlyLeuAsnTyrValValGlyLysMetGlu------------------Phe 229

Qy   859 GGAAGCGGGGACAGTGGCGTGATTGCAGCAACTGGCAAATGAGTTCCTTCCGTGTC 918
             |||||||||||          :||||||||||||||           ::|
Db   230 ArgLysSerGlnVal-----------GlnLysLeuAlaAspSerSerLeuGluIle 245

Qy   919 AATATAATATTCCTGAGCAACCCATGGCTATTTGAACGACCACTTATAAGCAGGTTTATGGAGAGA 978
             |||||||||||||||||||||||||          |||||         ||||:|||
Db   246 GlyAspVal------------------------------------------- 248

Qy   979 AATCCCTTAACCAATGCAATAATCAAGGAGCGACTCACCACAGGGACCCTCACCATATTCAAAGCAGG 1038
             ::||||||||:|||::          :|||||||||||||||||:
Db   249 ------------------ThrThrValAsnLeuThrGlnLeuArgGlyGly 259

Qy  1039 GTCAAGTTCAAATGCATCCCCCAGTGGCCCCAGGCCCACAGTGCTCAACTTCCGGATT 1092
             |::::: ::|||     |||||::||::  ||:||||     ||:::|:::|||||
Db   260 ValGlnSerAsnValValProProLeuGluGluAlaValProPheAspPheArgIle 277
```

```
RESULT 15
ABB47976
ID   ABB47976 standard; Protein; 379 AA.
XX
AC   ABB47976;
XX
DT   05-FEB-2002 (first entry)
XX
DE   Listeria monocytogenes protein #680.
XX
KW   Antibacterial; gene therapy; vaccine; biosynthesis; biodegradation;
KW   vitamin B12; bacterial infection; disease.
XX
OS   Listeria monocytogenes.
XX
PN   WO200177335-A2.
XX
PD   18-OCT-2001.
XX
PF   11-APR-2001; 2001WO-FR01118.
XX
PR   11-APR-2000; 2000FR-0004629.
XX
PA   (INSP ) INST PASTEUR.
XX
PI   Buchrieser C, Frangeul L, Couve E, Rusniok C, Fsihi H, Dehoux P;
PI   Dussurget O, Chetouani F, Nedjari H, Glaser P, Kunst F, Cossart P;
PI   Daniels J, Goebel W, Kreft J, Kuhn M, Ng E, Vazquez-Boland JA;
PI   Dominguez-Bernal G, Garrido-Garcia P, Tierrez-Martinez A, Amend A;
PI   Chakraborty T, Domann E, Hain T, Berche P, Charbit A, Durant L;
PI   Perez-Diaz J, Baquero F, Garcia Del Portillo F, Gomez-Lopez N;
PI   Maduenio E, De Pablos B, Wehland J, Kaerst U, Entian K, Hauf J;
PI   Rose M, Voss H;
XX
DR   WPI; 2002-010914/01.
XX
PT   Genomic sequence for Listeria monocytogenes, useful e.g. for treatment
PT   and prevention of Listeria and related bacterial infections, and
PT   related polypeptides -
XX
PS   Claim 6; SEQ ID NO 681; 192pp; French.
XX
CC   The present invention relates to the genome sequence of Listeria
CC   monocytogenes EGD-e (see ABA03041). The genome sequence and fragments of
CC   it are useful for selecting probes and primers for detecting genes in L.
CC   monocytogenes and related organisms, and for studying genetic
CC   polymorphisms and other genomes. The present sequence is a protein
CC   encoded by the genome sequence of the present invention. Proteins
```

```
CC   expressed from the genome sequence are useful for raising specific
CC   antibodies, identification of L. monocytogenes and related organisms, and
CC   for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC   B12. The genome sequence and proteins encoded by it are also useful for
CC   selecting compounds that regulate gene expression and cell replication
CC   and modulate L. monocytogenes-related diseases. In addition, the genome
CC   sequence and proteins encoded by it are useful in pharmaceutical and
CC   vaccines compositions for the treatment or prevention of infections by L.
CC   monocytogenes and related organisms.
CC   Note:The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence    379 AA;
```

Alignment Scores:
```
Pred. No.:            1.5e-10
Score:               203.00
Percent Similarity:   37.898
Best Local Similarity: 22.78%
Query Match:           7.35%
DB:                    23
```

```
                      Length:         379
                      Matches:         95
                      Conservative:    63
                      Mismatches:     155
                      Indels:         104
                      Gaps:            17
```

US-10-014-896-1 (1-1509) x ABB47976 (1-379)

```
Qy    226 CTGGCTGAGTTCGGAAAATACAATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTT 285
          ||||||||||| |||:::  ||  |||||||     |||||||||||||||||||||||||
Db     34 LeuAlaGluHisGly---------------------IleGluSerGluLys 43

Qy    286 ATCCACCATGAAGTCGTGGGAGAGAGTATAGCCACCGTTCACTATCCAAGGCTCCGACCCC 345
          :::||||||||||::|||  :::||| ||  |||||||||||||||||:||||||:::
Db     44 ValGlnTyrAsp---LeuAspArgAlaSerLeuValSerGluIleGlySerSerAsnGlu 62

Qy    346 AGCTTGCAGCCCTACCTGCTCAATGGCCTCACTTTGATGGTGGTCGCTCCCCCCTGAA---GAA 402
          ||||||||||||    |||||||   ||||||||||| |||    :||
Db     63 LysValLeuAlaPhe------SerGlyHisMetAspValValAlaAspAlaGlyAspValSer 80

Qy    403 GGCTGGGAGGTGCCCCCCAATTCTTCTGGGGTTGGAGCCGTGCATGGTGGTCGGGGC 462
          ||||||||    |||||   :: ||| ||||||||||||||||||||||||||||||
Db     81 LysTrpLysPheProProPheGluAlaThrGluHisGlyLeuGlyGluLeuGlyValAla 100

Qy    463 ACACTGGACAGAAGAACTCTGTGAGGGAGCCGATTACTGCGACCCGCCTGCTGATC 522
          |||       ||||| |||||:::|||  ||| |||||||||||||||||||||||
Db    101 AlaThrAspMetLysSerGlyLysGlyTyrAlaAspSerLeuAspGlyLeuIleIleGlyIle--- 115

Qy    523 AGGAAGTACATCCCCGAAGATCTTTCTCATTTCTCTGGGCCATGAATGAGGAGTCATCA 582
          |||||||| ||| ||||||||||||| ||| |||::||||||||:
Db    116 HisGluLeu---IleGluLeu---HisGluGlnLysGlnLys 124

Qy    583 ------GGGACAGGGGCTCAGAGGAATCTCAGCGCCCTGCTACAG 618
          :: ||| ||||||||||||| ||| :||| ||| ::||
Db    125 LeuAsnGlyLysIleArgLeuLeuAlaThrValGlyValIleGlyGluLeuGlyAla 144

Qy    619 TCAAGGGGCGTCCACTAGCCTTTCATTGTGGAGACGAGGGCGTTCATCGTTGGATGATTTC 678
          ||| :: :||||||:::||| ||| ||| |||||:::  |||
Db    145 GluGlnLeuThrGlnLysGlyLyTyrAlaAspGlyLeuIleIleGlyIle--- 163

Qy    679 ATTCCTAAACTTCAAGAAGCCCATCGCCTTGATTGCAGTCTTCAGAGAAGGGTTCCATGAAC 738
          |||||||||||| |||||||||| |||||||||||||||||||||||||||||||
Db    164 --ProSerGlyHisArg------IleValTyrAlaHisLeuLysGlySerIleAsn 178

Qy    739 CTCATGCTGCAAGTAAACAATGACTTTCAGGCCACTCGTTCCTCCCAAAGGAGACAAGC 798
          ::::||| |||||||||:: |||||||||||  |||  |||:::
Db    179 TyrThrValLysSerThrGlyLysAsnAlaHisSerSerMetPro------ 193

Qy    799 ATTGGCATCCTTTGCACTGCTGTCAGCCGATTGGAGCAGACACCAAATGCCTATATATTT 858
          ||||||||||||| || ::: ||| |||  ||||||||
Db    193 ------ 193

Qy    859 GGAACGCGGGACAGTGGTGCTGTATTGCAGCAACTGGCAAATGAGTTCCCTTCCCTGTC 918
          |||||||||| |||  ||| |||||||||||||||||||||||:::
Db    194 ------GluPheGlyVal 197
```

```
Qy    919 AATATAAATCCTGAGCAAACCCATGGGTTGTATTTTGAACCACTTTATAAAGGAGGTTTATTGGAGAGA 978
          |||  || |||||||||||||||:::   ||  :: ::|||||||||:::  ||||||||||||||||
Db    198 AsnAlaIle---AspAsnLeuLeuLeuPheTyrAsnGlyValGluLysPheValLysSer 216

Qy    979 AATCCCTTAACCAATGCAATAATCAACGGACC---ACCACGGCCACTCCACCATATTCAAAGCA 1035
          ||||||||  ||||| |||||||||||||||||   ::  :||||| ||||:::
Db    217 IleAspProAlaThrAsnGlnIleLeuLeuGlyAspPheIleHisAsnValThrValIleAspGly 236

Qy   1036 GGGGTCAAGTTCAAATGTCATCCCCCAGTGGCCAGCCCAGAGTCAACTTCCGGAATTCAC 1095
          |||      |||   ||:::|||||||||||||||||||||||||||||||||:::
Db    237 GlyAsnGlnValAsnSerIleProGluGlyLysAlaGlnLeuGlnGlyAsnIleArgSerIle 256

Qy   1096 CCT------GGACAGAGCAGTCCAAAGAGGTCCTA---GAACTCACGAAGAACATTGTG 1143
          |||   :::|||||||||| ||||||:: |||||  :::   ::
Db    257 ProGluMetAspAsnGluThrValLysGlnValLeuValLysIleIleLeuAsnLysLeuAsn 276

Qy   1144 GCTGATAAACACAGAGTCCAGTTCCATGTTGAGTGCCTTTGACCCCTCCCCGTC----- 1197
          |||||| |||||||||||||||||||||||||||||||||||||||:::
Db    277 LysGlnGluAsnValAsnLeuGluLeuIlePheAspTyrAspLysGlnProValPheSer 296

Qy   1198 ---AGCCCTTCTGATGACAAGACCCTTGGGCTACCAGCTGTCCGCGACCGTACAGTCC 1254
          :::  ||||||||||||||||||||||||||||||||||||||||||||||||
Db    297 AspLysAsnSerAspLeuValHisIleAlaLysSerValAlaSerAspIleValLysGlu 316

Qy   1255 GTCTCCCCGGAAGTCAAATATTACTGCCCCAGTTACTTCTTATTGGCAACAACAGCAGCCCA 1314
          |||  |||  |||||||||||||||:::   :::   |||:: ||||||||||||||||
Db    317 GluIleProLeuLeuGlyTyrIleSer------GlyThrThrAspAlaAla 330

Qy   1315 TTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTACCGAATTGATGATTTC 1374
          | |||||  ||||||||||||||:::||||||||||||||||||||||||||||
Db    331 GluPheThrLysAlaLys------LysGluPheProValIleLeuPheGlyPro 346

Qy   1375 GACTTCAAACGCATCCATGGAGTCAACGAGTCAACTCAGTCCAAGCCTAT 1425
          |||||||||||||||||||||:: ||||||||||||||||||||||:::
Db    347 GlyAsnGluThrProHisGlnValAsnGluAsnValSerIleGlyAsnTyr 363
```

Search completed: June 27, 2003, 10:57:55
Job time : 101.5 secs

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 27, 2003, 10:53:53 ; Search time 60.5 Seconds
(without alignments)
4795.598 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 2762
Sequence: 1 atggctaagcggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table:
BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 566448
Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104414_10394/app_query.fasta_1.1671
-DB=PIR_73 -QFMT=fastan -SUFFIX=n2p.rpr -MINMATCH-0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1 -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN-0 -MAXLEN-200000000000
-USER=US10014896_@CGN_1_1_80_@runat_27062003_104414_10394 -NCPU=6 -ICPU-3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS-1 -XGAPOP=10 -XGAPEXT-0.5 -FGAPOP=6
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP=6 -DELEXT=7

Database : PIR.73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	591	21.4	471	2	E87340	M20/M25/M40 family
2	540	19.6	576	2	S16693	Gly-X carboxypepti
3	53.5	19.2	596	2	T38349	carboxypeptidase s
4	361.5	13.1	443	2	T35974	probable aminoacyl
5	341.5	12.4	474	2	E87650	peptidase, M20/M25
6	330.5	12.0	467	2	B87070	probable peptidase
7	309.5	11.2	448	2	A70578	probable dapE2 pro
8	229.5	8.3	397	2	T19182	hypothetical prote
9	225.5	8.2	378	2	AB1469	succinyldiaminopin
10	223	8.1	410	2	H90212	succinyl-diaminopi
11	214.5	7.8	382	2	C97868	hypothetical prote
12	213.5	7.7	455	2	E71074	hypothetical prote
13	212	7.7	438	2	C96507	hypothetical prote
14	209	7.6	435	2	A96506	probable aminoacyl
15	207.5	7.5	399	2	T19180	hypothetical prote
16	205	7.4	383	2	G83500	succinyl-diaminopi
17	203.5	7.4	431	2	T29267	hypothetical prote
18	203	7.3	379	2	AB1108	succinyldiaminopim
19	202	7.3	407	2	JN0584	aminoacylase (EC 3
20	199	7.2	410	2	A64357	succinyl-diaminopi
21	197.5	7.2	383	2	A71650	succinyl-diaminopi
22	197	7.1	374	2	D95388	probable acetylorn
23	196.5	7.1	394	2	T19181	hypothetical prote
24	196.5	7.1	407	2	E89991	hypothetical prote
25	195.5	7.1	377	2	C82113	succinyl-diaminopi
26	195	7.0	388	2	C95017	succinyl-diaminopi
27	195	7.1	457	2	D97890	succinyl-diaminopi
28	194	7.0	388	2	C71961	succinyl-diaminopi
29	190.5	6.9	753	2	D64546	hypothetical prote
30	190.5	6.9	753	2	T05649	hypothetical prote
31	189	6.8	474	2	F75133	hypothetical prote
32	188	6.8	395	2	AB3543	succinyl-diaminopi
33	187.5	6.8	411	2	B71451	succinyl-diaminopi
34	187	6.8	384	2	B82973	probable desuccini
35	185.5	6.7	408	2	A47488	probable peptidic
36	185	6.7	381	2	F81073	aminoacylase (EC 3
37	184.5	6.7	398	2	B97404	succinyl-diaminopi
38	182.5	6.6	398	2	AB2622	succinyl-diaminopi
39	182.5	6.6	381	2	B81797	succinyl-diaminopi
40	181.5	6.6	381	2	B82846	succinyl-diaminopi
41	181	6.6	377	2	H72224	succinyl-diaminopi
42	180.5	6.5	396	2	B81307	hypothetical prote
43	180.5	6.5	365	2	B84940	succinyl-diaminopi
44	180	6.5	375	2	F84940	succinyl-diaminopi
45	179	6.5	403	2	H69362	succinyl-diaminopi
45	172.5	6.2	459	2	E75324	ArgE/DapE/Acyl fam

ALIGNMENTS

RESULT 1
E87340
M20/M25/M40 family peptidase [imported] - Caulobacter crescentus
C;Species: Caulobacter crescentus
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession: E87340
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87340
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-471 <STO>
A;Cross-references: GB:AE005673; NID:g13421965; PIDN:AAK22721.1; GSPDB:GN00148
C;Genetics:
A;Gene: CC0736

Alignment Scores:
Pred. No.:	2.1e-37	Length:	471
Score:	591.00	Matches:	151
Percent Similarity:	52.21%	Conservative:	85
Percent Identity:	33.41%	Mismatches:	184
Query Match:	21.40%	Indels:	32
DB:	2	Gaps:	12

US-10-014-896-1 (1-1509) x E87340 (1-471)

```
Qy  157 GAGGCGCTGAAAAGGTGCCATCCAGATTGCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAT 216
            |||  |||  ||||||||                   :::
Db   34 GluHisLeuAlaGluAlaIleArgPheGlnThrIleSerHisGlnAsnArgAlaAspAsp   53

Qy  217 ACTACAAGCCCTGGCTGAGTTCGGAAAA---------TACAATTCATAAAGTCTTTCCTACA 267
            ||||||||          |||               :::::::
Db   54 GlnPro------AlaGluTrpAspLysLeuHisAlaTrpLeuGlnThrThrTyrProGln   71
```

```
Qy  268  GTGGTCAGCACAGCTTTATCCAGCAATGAAGAGTCGTGGAAGAGTATAGCCACCTGTTTCACT  327
              |||||        ::              ||||            :||||||||    |||
Db   72  ------AlaHisLysAlaMetThrArgGluValValAlaGlyHisGlyLeuValTyrThr  89

Qy  328  ATCCAAAGCTCCGACCCCAGCGTTGCAGCCCCTACCTGTGATGGGTCACTTTGATGTGGGTG  387
              ||||||||            ::::||||||||||||||        |||
Db   90  AspGlyLysValTrpGlyLysArgGlyIleAlaIleAspAspSerLeuValThrIlePhe  149
         TrpThrGlySerAsnProAlaLeuAlaProIleValIleAsnThrHisSerAspValVal  109

Qy  388  ------CCTGCCCCTGAAGAGACGGGATCGCCCCCCATATTCTCATGGGCGTTGGAGCGT  438
                   |||                      |||
Db  110  ProValThrProGlySerGluGlyGlnTrpThrHisProProPheAlaGlyValValAla  129

Qy  439  GATGGCGGTCATCTATGGTGGTCGGGGGCACACTGGACGACGAAGAAACTCTGTGANGGATTACTG  498
              ||||||            :||||||||||
Db  170  SerGlyHisAspGluGlnValAlaArgGlyValGluGlyAlaGlnAlaLeuAlaAlaLeuLeuLys  189

Qy  499  CAGGCCTGGAGCGTCTGCGTGATCAGGAGAGTACATCCCCAGCAGAGAGCTTTCTTCATTTCT  558
              |||||||||||                       :||
Db  190  SerArgAsnIleLysAlaGlnInPheValLeuAspGluMetAlaValAlaAlaAspHis  209

Qy  559  CTGGGGCCANGATGAGGAGTCATCAGGGAGCACAGGGGGCTCAGAGGGATCTCAGCCTGCTACAG  618
              :::::::|||||||||||                    :||||
Db  210  ---ProValThrArgGluProAlaIleAlaIleGlyValAlaLeuGlyTyrAlaThr  228

Qy  619  TCAAGGGGCGTCCAGCCGTTCAGTGTGACGAGCGGGGGGCTTCANCTTTGGATGANTTTC  678
              ||||||||                    :::||||||||||||||
Db  229  LeuLysValThrAlaProAlaValGlyHisSerSerAlaProProLysAspGlyGly  248

Qy  679  ATTCCTAACTTCAAGAAGCCCATCGCCTGATTGCAGTCTCAGAGAAGGGTTCCANTGAAC  738
              :::||||||          :::||||||||||||||
Db  249  GlyValValThrLeuSerLysValAlaValGlnAlnIleHisAspAsnProPheProMetLys  268

Qy  739  CTCATGCTGCAAGTAAACATGACTTCAGGGCCACTCTTCAGCTCTTCCACCAAAGGAG---ACA  795
              ::::||||                    :||||||
Db  269  Phe---GlnGlyProGlyAlaAspMetLeuLysValAlaIleSerProHisAlaSerProVal  287

Qy  796  AGCATTGGGCATCCTTGCACTGCTGTCAGCCGATTGGAGCGAGGACACCAAATGCCTATCATA  855
              ::::|||                    |||  :::|||
Db  288  ValLysValPheAlaAlaAsnThrTrpLeuPheSerSerLeuLeuValLysValThrAla  307

Qy  856  TTTGGAACGGGGACAGTGGTGACTGTATTGCAGCAACTGCAAATGAGTTTCCCTTCCCT  915
              |||                    :::||||||||||||||||
Db  308  LysSerProAlaGlyLyAlaAlaMetLeuHisThrThrIleAlaProThrMetLeuLysGly  327

Qy  916  GTCAATAATAATCCTGAGCAACCCATGGGATTGGACCACTTTAAGGCAGCGTTTTAYGGAG  975
              |||||||||                    :||
Db  328  SerProLysGluAsnValLeuProGlnAsnAspAlaThrArpIleAsnTyrArgIleAla  347

Qy  976  AGAAATCCCTTAACCAAATGCAAATAAATCAGGAACCCACCGACTCACCATAATCAAAGCA  1035
              :::::::|||                    :::
Db  348  ProGlyAspSerSerSerAlaLysGluAlaValGlyAspLeuPro  367

Qy  1036  GGGGTGTCAAGTTCAATGTCATCCCCAGTGGCCCCAGGCCCACAGTCAACTTCCGGAATTCAC  1095
               |||||||                    :::|||
Db  368  ValGluLeu------AlaPheGluGlyHisArgAsnGluProSerAlaValSer  383

Qy  1096  CCTGGACACAGTCCAAGAGTCCTAGAAACTCACGAAGAACATTGTGGCCGATAACAGA  1155
               ||||||                    |||
Db  384  SerThrThrSerAspAlatrpLysThrLeuAlaGlyLeuAlaAlaAspGluSerGlnAla  403

Qy  1156  GTTCCAGTTCCATGTGTTGGAGTGCCTT------GACCCCTCCCCGTCAGC  1200
               |||||||||||                    ::|||
Db  404  ProValValProGlyLeu---ValThrAla  416

Qy  1201  CCTTCTGTGATGAC------AAGGCCTTGGGGCTACAGCTGCTCCGCCAGCCGTACAG  1251
               |||||||                    |||
Db         GluLysLeuSerAsnAlaValAlaArgIleProThrVal---GlyThrAspSer  416

Qy  1252  TCCGTCTTCCGGAAGTCAAATATACTGCCCCAGTACTTCTTATTGGCAACACAGACAGC  1311

Qy  1312  CGAATTCTTTACAAACCTACCACTGGCATACACAGGTTGTCACCCATCTACATAACAGCT  1371
```

RESULT 2

```
S16693
Gly-X carboxypeptidase (EC 3.4.17.4) precursor - yeast (Saccharomyces cerevisiae)
N;Alternate names: carboxypeptidase yscS; protein J0510; protein YJL172w
C;Species: Saccharomyces cerevisiae
C;Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 29-Oct-1999
C;Accession: S16693; S16881; S56955; S53878
R;Bordallo, J.; Bordallo, C.; Gascon, S.; Suarez-Rendueles, P.
FEBS Lett. 283, 27-32, 1991
A;Title: Molecular cloning and sequencing of genomic DNA encoding yeast vacuolar carb
A;Reference number: S16693; MUID:91243870; PMID:1709881
A;Accession: S16693
A;Molecule type: DNA
A;Residues: 1-576 <BOR>
R;Sporman, D.O.; Heim, J.; Wolf, D.H.
Eur. J. Biochem. 197, 399-405, 1991
A;Title: Carboxypeptidase yscS: gene structure and function of the vacuolar enzyme.
A;Reference number: S16881; MUID:91224132; PMID:2026161
A;Accession: S16881
A;Molecule type: DNA
A;Residues: 1-386, 'T', 388-576 <SPO>
A;Cross-references: EMBL:X57316; NID:g35593; PIDN:CAA40571.1; PID:g35594
R;Obermaier, B.; Piravandi, E.; Rinke, M.; Domdey, H.
submitted to the Protein Sequence Database, September 1995
A;Reference number: S56937
A;Accession: S56955
A;Molecule type: DNA
A;Residues: 1-576 <OBE>
A;Cross-references: EMBL:Z49447; NID:g1008366; PIDN:CAA89467.1; PID:g1008367; MIPS:YJ
R;Bordallo, J.; Suarez-Rendueles, P.
Mol. Gen. Genet. 246, 580-589, 1995
A;Title: Cis and trans-acting regulatory elements required for regulation of the CPS1
A;Reference number: S53878
A;Accession: S53878
A;Molecule type: DNA
A;Residues: 1-78 <BOW>
C;Genetics:
A;Gene: SGD:CPS1
A;Cross-references: SGD:S0003708; MIPS:YJL172w
A;Map position: 10L
C;Function:
C;Description: hydrolase; metallo-carboxypeptidase
C;Keywords: glycoprotein; hydrolase; metallo-carboxypeptidase; yeast vacuole
F;1-20/Domain: signal sequence #status predicted <SIG>
F;21-576/Product: Gly-X carboxypeptidase #status predicted <MAT>
F;88,176,228,381,525/Binding site: carbohydrate (Asn) (covalent) #status predicted
```

Alignment Scores:

Pred. No.:		
Score:	1.89e-33	
	540.00	
Percent Similarity:	47.49%	
Best Local Similarity:	31.86%	

	Length:	576
	Matches:	159
	Conservative:	78
	Mismatches:	184
	Indels:	78
Query Match:	19.55%	
DB:	Gaps:	19

US-10-014-896-1 (1-1509) x S16693 (1-576)

```
Qy  157  GAGGCGCTGAAAGGTGCCATCCAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAT  216
              |||    |||                    :::||||||
Db   92  GluLysLeuSerAsnAlaValAlaArgIleProThrVal-----GlnAspLysAsnPro  109
```

```
Db  417  ArgTyrMetGlyGlyValSerSerAspValTyrArgPheGlnProLeuValLeuThrVal  436
              ||||::                    :::||||||||    :::
Qy  1372  GAAGACTTCAAACGCCATCCATGGAGTCAACAGGAGAAAAATCTCAGTCCAAGCCTATGAGACC  1431
Db  437  AspGlyThrLysValIleHisGlyThrHisAspGluHisIleSerSerLeuAspAsnValGluArg  456

Qy  1432  CAAGTGAAATTCATCTTTGAGTTGATTCAAGAATGCT  1467
               |||::||||||                    |||
Db  457  MetValArgPheTyrGlnArgLeuValGluThrAla  468
```

Qy   217 ACTACAGCC--------------------------CTGGCTGAGTTCGGAAAATACATTCAT 252
           ||| |||
Db   ...

Qy   253 AAAGTTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTAT 312
Db   110 AsnProAlaAlaAspProAspProAspPheTyrLysHisPheTyrLysHisAspTyrProGlu 129

Qy   313 AGCCACCTGTTCACTATCCAAGGCCTCGGAGACCCCAGCTTGCAGCCCTACCTGCTGATGGCT 372
Db   148 GlyLeuLeuThrTyrThrTrpGluGlyTyrSerAspProAspProLeuLeuLeuMetAla 167

Qy   373 CACTTTGATGTGGGTGCCTGCCCTGCCCCTGAAGAA--------GGCTGGAGGTGGCCCCATTC 423

Qy   424 TCTGGG------------------TTGGACGGTGATGGCGTCGGGGGCGCACACTGACGAC 474
Db   188 SerGlyHisTyrAspProGluThrAsp---PheValTrpGlyArgGlySerAsnAspCys 206

Qy   475 AAGAACTCTGTGATGGCCATTACTGACGGGCTGCTCCTGCCTTGCGATCAGGAAGTACATC 534
Db   207 LysAsnLeuLeuIleAlaIleLeuGluLeuLeuIleLeuAspGlyPheLys 226

Qy   535 CCCCGGAAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGACTGGATCATCAGGGACA---GGG 591
Db   227 ProAsnArgThrIleValMetSerLeuGlyPheAspGluGluAlaSerGlyThrLeuGly 246

Qy   592 GCTCAGAGGATCTCAGCCCCTGCTACAGTCAAGG--------GGCGTCCAGCTAGCC 639

Qy   640 TTTCATTGTGGACGAGCGGGGGGCGTTCAATCTTG----GATGATTTCATTCCTAAC 687
Db   265 SerIleIleLeuGluGlyLysSerValPheValAlaThr 284

Qy   688 TTCAAGAAGCCCATCGGCCTTGATTGCAGTGTTCCAGAAGAGGGTTCCATGAACCTCATGCTG 747
Db   ...:::||||||| ProLeuAsn AlaGluLysGlyTyrValAspPheGluVal 297

Qy   748 CAAGTAAACATGACTTCAGGCCGCCACTCTTCCAAAGGACAGACAAGCACATTGGCATC 807

Qy   808 CTTGCAGCCTGCTGTCGTCCCGATTGGAGCAGACACCAATGCCTATCATATTTGGAACGGGG 867
Db   298 SerIleLeuGlyLyHisGlyHisSerSerValProProAspHisThrThrIleGlyIle 317

Qy   808 AlaSerGluLeuIleThrGluProHeGluValAlaAsnProPheAspThrTyrGluPheGluAsp 337

Qy   868 ACAGTGTTGGTACTGTTATGCAGCAAACTGGCAAATGAGTTTCCCTTCCCTGTCAATATAATC 927

Qy   338 AsnProIleTyrGlyLeuLeuThrCysAlaAlaLeuGluHisSerLysSerLysLysAsp 357

US-10-014-896-1 (1-1509) x T38349 (1-596)

Qy   28   GCCCTGGTGGCTATGCGTTGGCTCCCTAGTTTGTCCCTACCGTCTCCGAATCGATGGGCCCGAGG 87
           |||||||| |||::|||
Db   40   AlaPheValGlyPheGlnPheLeuLeuIleLeuIlePheValThrTyrPheLeuLeuArgGlyGlySerAsn 59

Qy   88   ACCGGGGAGCATCAAAGGCGTCGCGGAATCCCTTCTCAGTCAAAGAGGAAACCGGTC 147
Db   60   AspAsnAspLysGlnGluMetSerHisSerProGlySerCysMetAspSerCysSerGluSerAla 79

Qy   148 GCCATGAAAGAGGCGCTGAAA-------------------------- 168
Db   80   AlaValSerThrSerAlaLysCysTyrIleProProValLeuThrProAlaLysGluPro 99

Qy   168 -------------------------- 168

Qy   167 -------- 167

RESULT 3
T38349
carboxypeptidase s precursor - fission yeast (Schizosaccharomyces pombe)
C;Species: Schizosaccharomyces pombe
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C;Accession: T38349
R;Murphy, L.; Harris, D.; Barrell, B.G.; Rajandream, M.A.; Wood, V.
submitted to the EMBL Data Library, August 1995
A;Reference number: Z21787
A;Accession: T38349
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-596 <MUR>
A;Cross-references: EMBL: Z98601; PIDN:CAB11265.1; GSPDB:GN000066; SPDB:SPAC24C9.08
A;Experimental source: strain 972h-; cosmid c24C9
C;Genetics:
A;Gene: SPDB:SPAC24C9.08
A;Map position: 1

Alignment Scores:
Pred. No.:   8.65e-33
Score: 531.50
Percent Similarity: 40.75%   Matches: 149
Best Local Similarity: 25.308   Conservative: 91
Query Match: 19.24%   Mismatches: 212
DB: 2   Indels: 137
                                          Gaps: 13
                                          Length: 596

Db   458 GluThrGluLeuGlyHisIleAspIleThrLeuLeuGluArgGluGluProAlaProLeu 477
Db   478 SerProSerSerGlyProVal------TrpAspIleLeuAlaGlyThrIleGlnAspVal 495

Db   516 ---GlyAsnThrAspThrThrLysTyrTyrPheAsnLeuSerLysAsnIleGluTyrArgPheVal 534
Db   535 GlySerIleIleLeuGluAspLeuLeuLysThrLeuHisSerValAsnGluHisValAsp 554
Db   555 ValProGlyHisLeuSerArgAlaIleLeuAlaPheValTyrGluTyrTyrIleValAlaAsnValAsn 573

Db   438 LysIleLeuAlaGluLeuLysTyrGlyTyrGlyLeuSerLysSerAsnGlyAspAspProTyrIleIlePro 457
Db   120 SerLeuValArgPheGlnLeuLeuLeuArgIleLeuProThrValCysTyrAspMetGly 139

```
Db   140 AspValGlyAspAspArgPheAspIlePheAlaValPheGlnAspLysValArgGlu 159
         :::
Qy   256 GTCTTTCCTACAGTGGTCAGCACCAGCTTTAYCCAGGATGAAGTCGTGGAAGAGTAYAGC 315
         :::::|| |||
Db   160 LeuTyrProAsnIlePheLysLys------LeuLysValGluTyrValAsnThrTyrGly 177
         |||||||||||
Qy   316 CACCTGTTCACATCACATCAAAGGCTCGGGACCCCAGCTTGCAGCCCTACCTGATGGCTCAC 375
         |||
Db   178 LeuLeuIleThrLeuGluGlySerAsnLysAspLeuLysProLeuValLeuMetGlyHis 197
         ::: ||||| ||
Qy   376 TTTGAYCTGGGTGCCT------GCCCTGAAGAGGCTGGGAGGTGCCCCAGTTCTCT 426
         |||||
Db   198 GlnAspValProValAsnGlnAlaSerLeuAspArgTrpTyrPheProProPheSer 217
         ::: |||
Qy   427 GGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGGCACACTGGACACAGAAACTGTG 486
         ||||||||| |||
Db   218 AlaThrTyrHisAsnGlyHisValTyrSerArgGlyAlaAlaAlaAspLysAsnSerVal 237
         ::::: ||| |||
Qy   487 ATGGCAATTACTGCAGGCCTTGGAGCTCCTGATCAGGAAGAGTACATCCCCGAAGATCT 546
         ||||||||||||
Db   238 ValAlaIleLeuGluAlaLeuGluIleIleSerAspTyrLysProGluGlnThr 257
         || |||
Qy   547 TTCTTCAATTCTCTGGGCCCATGAGTGAGGAGTCATCAGGG---ACAGGGCTCAGGAGGATC 603
         ||| |||
Db   258 ValIleLeuAlaSerPheGlyPheAspGluGluValSerGlyTyrArgGlyAlaLeuProLeu 277
         ||| :::
Qy   604 TCAGCCCTGCTACAGTCAAGG------GGCGTCCAGCTACCTTCCATTGTTGGAC 651
         ||| :::
Db   278 AlaHisLeuTyrGluArgTyrGlyLysAspGlyVal------AlaLeuIleLeuAsp 295
         ||||||| |||::.:||
Qy   652 GAGGGGGGCTTCATCATGTTGGADGATGATCATTCATCCTAACTTCAAGAGAGCCCAATCGCCTTGATT 711
         ||||||||||||
Db   296 GluGlyGlyPheThrIleAsnLeuPheGlyThrLeuPhe------AlaThrVal 311
         |||
Qy   712 GCAGGTCTCAGAAGAGGGTTCCATGAAACCATGACCTGCAAGTAAACATGACTTCAGGCCAC 771
         |||
Db   312 CysValAlaGluLysGlyTyrMetAspValHisLeuLeuLysThrProGlyGlyHis 331
         ||| :::::::
Qy   772 TCTTCAGCTCCTCCAAAGGAGGACAAGCAACCATTGGCATCCTTGCAGCTGTCTGTCAGCCGATTG 831
         ||||||
Db   332 AlaSerIleProProProHisThrAsnIleGlyLeuMetSerLysLeuValThrGlnIle 351
         |||:::::::||| :::
Qy   832 GAGCAGCACCAAATGCCTAYCATATTGGAAGCGGGACAGTGGTCGTTATTCGAGCAA 891
         |||
Db   352 GluGlu------ 353
         |||
Qy   892 CTGGCAAATGAGTTTCCCTTCCCGTCCAATATAATCCTGAGCAACCATGGCTATTTGAA 951
         :::
Db   354 ------ProPheGlyGlyGluLeuThrPheGluAsnPro------PheTyr 366
         |||
Qy   952 CCACTTATAAGCAGGTTTATGGAG--- 975
         ||| |||
Db   367 ThrThrIleuGlnCysPheAlaGlyAlaAsnSerAlaAspMetAspSerAspAsnLeuArgGlnLeu 386
         |||
Qy   976 ------AGAAATCCCTTAACCAAT 993
         |||
Db   387 IleLysSerGlyAspThrGlyLysMetThrAspLeuPheSerLysSerArgLeuThrArg 406
         ||||||| :::
Qy   994 GCAATAAATCAGGACCACCAGCCACTCACCATATTCAAAGCAGGGTCAGTTCAAATGTC 1053
         ||||||||| :::
Db   407 TyrPhePheGluThrHrSerIleAlaValLeuAsnGlyValValLysValAsnAla 426
         |||
Qy  1054 ATCCCCCACAGTGGGCCCACAGTCAACTTCCGGACAGACAGTCCAA 1113
         ::::|||
Db   427 LeuProGluGluThrThrLeuGluAlaValAlaAsnHisArgValAspAlaLeuArgLysGlyLeuLys 446
         ||| :::
Qy  1114 GAGGTCCTAGAAACTTCACGAAGAGAACACATTGTGGCTGCGATAACAGATGCCAGTTCAGTGC--- 1170
         :::::|||
Db   447 GlnValTyrAspArgTyrGlyGlyLeuLeuGluGluGluGlyIleuPheGlyHisGluTyrHisValAsn 466
         ||| :::
Qy  1170 ----------- 1170
```

```
Db   467 ValThrLeuPheAsnGlyLeuThrHrValValGlyGluThrGluAspAlaIleGlyHisIlePhe 486
         ||| :::
Qy  1171 ---------TTGAGGTGCCTTTGACCCCCTGCCCCCCGGGTCAGCCCCTTCTCGAGGAGCCTTG 1221
         ||||||
Db   487 AlaSerThrAlaLysThrLeuGluProSerProValSerProTyrAspGluSerSerAsp 506
         :::::||
Qy  1222 GGCTACCAGTCGCTCGGCCAGACCGTACAGTCGGTCGTTCCCGGAGGTCAATATTACTGCC 1281
         ::::: |||
Db   507 AlaTyrLysLysLeuValAlaGlyValAlaIleLeuArgTyrThrPheGlyThrSerValThr 526
         ::::: |||
Qy  1282 CCAGTTACTTCTATTGGCACAACAGCAGACAGCCGATTCTTTACAAACCTCACCACTGCATC 1341
         ||||||||| :::
Db   527 ProAlaLeuMetProAlaAsnThrAspThrHrArgHisTyrArgProAsnLeuThrSerAsnIle 546
         :::::|| |||
Qy  1342 TACAGGTTCTCACCCCATCTACATACAGCCGTGAAGACGTCAAAGCGAATCCATGGA------ 1395
         |||||::|| |||
Db   547 TyrArgGlyTrpThrProValSerThrAsnSerThrLysAsnSerPheAsnGlyHisThr 566
         ||||||||| :::
Qy  1396 GTCAACGAGAAAAATCTCAGTCCAAGCCCTATGAGACCCAAGTGAAATTCATCTTTGAGTG 1455
         :::::||||
Db   567 IleAsnGluAsnMetArgTyrAspAlaHisMetAspSerIleGluPhePheTyrAsnPhe 586
         ::::::||| :::
Qy  1456 ATTCAGAATGCTGACACAGACCAGGAG 1482
         |||
Db   587 IleLeuLeuValSerAspSerGlyGluGlu 595
         ||| :::
```

RESULT 4
T35974
probable aminoacylase - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C;Accession: T35974
R;Seeger, K.J.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M
submitted to the EMBL Data Library, January 1999
A;Reference number: 221551
A;Accession: T35974
A;Status: Preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-443 <SEE>
A;Cross-references: EMBL:AL035161; PIDN:CAA22735.1; GSPDB:GN00070; SCOEDB:SC9C7.23
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC9C7.23
C;Superfamily: succinyl-diaminopimelate desuccinylase

```
Alignment Scores:
Pred. No.:
Score:                   9.87e-20
Score:                   361.50      Length:        443
Percent Similarity:      45.63%      Matches:       134
Best Local Similarity:   29.26%      Conservative:  75
Query Match:             13.09%      Mismatches:    181
DB:                      2           Indels:        68
                                     Gaps:          20

US-10-014-896-1 (1-1509) x T35974 (1-443)

Qy   156 AGAGGCGCTGAAAGGTGCCATCCAGATTCCAACAGTGACTTTAGCTCTGAGAGTCCAA 215
         ||||||
Db    34 ArgGlyLeuGlyAspCysGln-GluArgProAlaLeuAlaLeuAlaAlaArg------ 51
Qy   216 TACTACAGCCCTGGCTGAGTTCGGAAAATACATTCATAAAGTCTTTCCTACAGTGGTGTCAG 275
         ||||||||
Db    52 ------LeuAlaGluAlaGly------IleGluProHrLeuLeuGl 63
Qy   276 CACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTGATAAGCCACCTGTTCACTATCCAAGG 335
         :::
Db    63 uArgThrAlaGlyArgThrAsnValValAlaArg--- 77
Qy   336 CTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTTGGGTGCCC 395
         |||
Db    77 yThrAspProSerAlaAspProAlaLeuLeuValHisGlyHisLeuAspValProAlaGl 97
Qy   396 TGAAGAAGGCTGGGAGGTGCCCCCATTCTCTCGGGTTGGAGGCGTGATGGCCGTGCATCGTCTATGG 455
         ::::::||
```

RESULT 5

E87650

peptidase, M20/M25/M40 family [imported] - Caulobacter crescentus

C;Species; Caulobacter crescentus
C;Date; 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C;Accession; E87650
R;Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A;Title: Complete Genome Sequence of Caulobacter crescentus.
A;Reference number: A87249; MUID:21173698; PMID:11259647
A;Accession: E87650
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-474 <STO>
A;Cross-references: GB:AE005673; NID:g13424925; PIDN:AAK25201.1; GSPDB:GN00148
A;Gene: CC3239

Alignment Scores:
Pred. No.:
Score: 341.50
Percent Similarity: 43.98%
Best Local Similarity: 27.35%
Query Match: 12.36%

Length: 474
Matches: 125
Conservative: 76
Mismatches: 235
Indels: 21
Gaps: 12

Pred. No.:   3.5e-18

US-10-014-896-1 (1-1509) x E87650 (1-474)

[Nucleotide / protein sequence alignment — Qy (query) vs Db (database) lines follow; positions indicated at left and right of each line.]

RESULT 6
B87070
probable peptidase [imported] - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 02-Aug-2002
C;Accession: B87070
R;Cole, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.; Ho
R; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holroyd,
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A;Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.; Sq
A;Title: Massive gene decay in the leprosy bacillus.
A;Reference number: A86909; MUID:21128732; PMID:11234002
A;Accession: B87070
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-467 <STO>
A;Cross-references: GB:AL450380; NID:g13093213; PIDN:CAC331669.1; GSPDB:GN00147
C;Genetics:
A;Gene: ML1288
C;Superfamily: succinyl-diaminopimelate desuccinylase

Alignment Scores:

Pred. No.:		
Score:	2.45e-17	Length: 467
Percent Similarity:	46.21%	Matches: 114
Best Local Similarity:	28.79%	Conservative: 69
Query Match:	11.97%	Mismatches: 166
DB:		Indels: 47
Query Match:	2	Gaps: 17

us-10-014-896-1 (1-1509) x B87070 (1-467)

```
Qy   219 GluValGlnAlaGlyGlyLysValTyrGlnAspPheThrLeuThrValThrAsnProGly 238

Qy   766 GGCCACTCTTCAGCTCCTCCAAAAGGAGACAAAGCAATTGGCATCCTTGCAGCTGCGTCAGC 825
         |||||||||
Db   328 ATCCAAGGCTCGGACCCCAGCTTGCGAGCCCAGCCCTCGATGGGCTCACTTTGAGTGGTG 387
         :::    |||  |||:::    |||          |||  |||||||
Qy   239 GlyHisSerSerArgProValSerProAsnAlaIleThrArgLeuSerAlaAlaLeuAsp 258
Db    93 LeuIleAlaGlyGluAspSerSerArgGlyAlaLeuLeuIleHisGlyHisLeuAspVal 112

Qy   937 CCATGCTATTTGAACCACACTTATAAGCAGTGTTATGGAGAGAAAATCCCTTAACACAATGCA 996
         |||||||
Db   388 CCTGCCCTGAAGAAAGGCCTGGGAGGTGCCCCAATTCTCTCCGGGTTGGAGGCGTGGGCGTC 447
         :::    |||  |||:::    |||
Qy   259 ArgIleGlyAlaTyrGlnPheProProArgPheAsnAspAlaThrArgGlyThrArgPheThr 278

Qy   826 CGGATTGGGAGCAGAGCACCAAGTCCTAACATAATTTGGAAGCGGGACAGTGGTGACTGTATTG 885
         |||:::
Db   113 ProAlaGluThrAlaGluTrpProSerValHisProPheSerGlyAlaValAlaValGlyGln 132

Qy   299 ------ValAsnAspProAlaIleLeuIleThrLysAspArgTrpProAsnSer 316
         :::    |||:::
Db   997 ATAAATCAGGACCACCACAGCACTCACCATAATTCAAAGCAGGGTGCAAGTTCAAATGTCATC 1056
         |||:::||||||||
Qy   448 ATCTPATGGTTCGCGGCCACACTGGACACGACCAAGAACTCTGTGAATGGCCATTACTGCAGGTTG 507

Qy   317 MetLeuArgThrThrCysValAlaIleThrMetValSerAlaGlyHisAlaProAsnAlaLeu 336
Db   317 MetLeuArgThrThrCysValAlaIleThrMetValSerAlaGlyHisAlaProAsnAlaLeu 336
         :::    |||||||    :::
Qy   133 ValIleTrpGlyTyrArgGlyTyrAlaIleAlaIleAspAspMetLysAspMetValGlyMet 152

Qy  1117 GTCCTAGAACTCACGAAGAAACATGTGGTGACTGATAAACAGAGATCCAGTTCCATGTGTTGAGT 1176
         |||||||
Db   508 GAGTCCTGCTGATCAGGAGGTAGTACATCCCCGAGAGATCTTTCTTCATTTCTCTGGGGCAT 567
         :::||||||:::
Qy   337 ProGlnArgAlaThrAlaAsnIleAsnCysArgIleLeuProGlyThrProIleAspGlu 356
Db   153 ArgGlnLeuGlyValAlaGlyIleLeuAlaProProArgAspProArgLeuValPheAla 172

Qy  1057 CCCCCAGTGGCCCAGGCCACACTTCCGGATTCACCCTGGACAGACAGTCCAAGAG 1116
Db   133 ValThrTrpGlyTyrArgGlyTyrAlaIleAlaIleAspAspMetLysAspMetValGlyMet 152

Qy   375 AlaHisSerLysProAlaSerProProLeuThrProAlaIleMetAlaPro 394
Db   568 GATGAGGAGTCACTCAGGGACA---GGGGCTCAGAGGAGTCAGCCCCGCTACAGTCAAGG 624
         :::    :::  |||
Db   190 ProAspLeuPheAspGlyValThrGluAla---IleGlyGluValGlyGlyPheSerLeu 208

Qy  1234 CTCCGCCACCGTACAGTCCGTCTCCCGGAAGTCAATATTACTGCCCCAGTTACTTCT 1293
         |||:::
Db   394 AlaHisSerLysProAlaSerProProLeuThrProAlaIleMetAlaPro

Qy  1294 ATTGGCAACACAGCAGCCGATTCTTTACAAACCTCACCACTGGCATCTTACCAGTTCTCTAC 1353
         |||
Db   670 GATGAGTTCAATTCCT------AACTTCAAGAAGCCCATGCCTATTGTAGCCACTTATA 720
         :::
Qy   414 ThrGlyAlaThrAspAlaValHisThrSerAlaAlaGlyIleLeuProThrTyrGlyValThr 433
Db   227 GlyLysGlyMetGlnTrpMetArgLeuThrArgAlaArgGlyAlaGlyHisGlySerMet 246

Qy  1354 CCCATCTACATACACACCCTGAAGAACTTCAAACGCCAACCATGGGAGTCAACGAGAAAATCTCA 1413
         |||||||
Db   721 GAGAAGGGGGTTCCCATGAACCTCATCGCTGCAGAGTAAACATGACTTCAGCTTTCAGCT 780
         :::  |||||||
Db   209 Thr------ValProCysArgAsnGlyValGluArgGlyValGlnTyrLeuIleGluThrAla 226

Qy   434 GlyLeuPheHisGlyProGlyGly---ThrGlyAlaHisGlyLeuAsnGluArgMetArg 452
Db   247 ValHisAsnGlnAsnAlaValThrAlaValAlaGluLeuAlaLysArgLeuGlyArgHis 266

Qy  1414 GTCCAAGCCTTATGAGACCCAAGTGAAAATTCATCTTTGAGTGATTCAGAAT 1464
         |||||||
Db   781 CCTCCAAAGGAGACAAAGCAAGATATGCCATCCTTGCAGCTGCGTCGAGCAGAGCACCAAGTC 840
         :::  |||
Db   448 ...

Qy   453 ValLysSerLeuThrYrGluGlyValArgAspPheLeuHisGlyLeuIleGlnAsp 469
Db   841 CCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAAT 900
         :::
Db   267 GlnPheProLeuVal---ThrThrAspProValValGlnLeuProPheLeuAlaIleSerGlu 285

Qy   901 GAGTTTCCCTTCCCTGTCAATATAATACCTGAGCAACCCATGGCTATTGTGGAACCACTTATA 960
         :::||||||
Db   286 GluThrGlyLeuGluPheAspIle------GlySerProAspLeu---GluGlyAlaIle 302

Qy   961 AGCAGGTTTATGGAGAGAGAAATCCCTTAACACAATGCCTATGCTTTAGTGGAGTT 1020
         :::
Db   303 GluLysGluGlyProMetAlaArgMetLeuLysAlaValLeuThrArgAspThrAlaAsnPro 322

Qy  1021 ACCATATTCAAACGCCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCACCACAGTC 1080
         ||||||
Db   323 ThrValLeuLysAlaGlyTyrArgLysValAsnValValProAlaThrAlaMetVal 342

Qy  1081 AACTTCCGGATTCACCCTGGACAG------------ACAGTCCAAGAGGTTCCTTA 1122
         |||
Db   343 AspCysArgIleLeuProGlyAsnGlnAlaAlaAlaPheGluAlaLeuAspIleIle 362

Qy  1123 ------GAACTCACGAAGAAACATTGTGGCTGATGATAACAGAGATCCAGTGTTGAGT 1176
         ||||||:::
Db   363 GlyProAspValThrArgGluTrpProIleLysAspLys 373

Qy  1177 GCCTTGACCCCCTCCCGGTTCACCCTTCGTGATGACAAGGCCCTTCGGGGCTACCAGCTCGCTC 1236
         :::
Db   374 ------LeuProProTyrGluThrThrAlaPheAspGlyAspLeuVal------AspAlaMet 389
```

```
Qy   1237 CGCCAGACCGTACAGTCCGGTCTCCGGAGGAAGTCAATATTACTGCCCAGTTACTTCTATT 1296
          |||    |||    ||||||    ::||||    ||||    |||    :||    |||
Db    390 AsnAlaAlaValLeuAlaLeuValAlaAspProAsp---GlyArgThrValProTyrMetAlaSer 408

Qy   1297 GGCAACACACAGCCGATTCTTTACAAACCTCACCACTGGCACTCACGGTTGTCACCCC 1356
          |||||||||||:|:||||    |||    |||    ||:    |||    |||
Db    409 GlyGlyThrAspAlaLysAlaArgLeuGlyIleArgCysPheGlyPheThrPro 428

Qy   1357 ATCACACATACAGCCTGAA---GACTTCAAACGCATC---CATGGAGTCAACGAGGAAAATC 1410
          :::    |||    |||    |||    ||:    |||    |||
Db    429 LeuArgLeuProProGluLeuAspPheThrAlaLeuPheHisGlyValAspGluArgVal 448

Qy   1411 TCAGTCCAAGCCTATGAGACCCAAGTGAAATTCATCTTTGAGTTGATT 1458
          ::::|||||||    |||    |||    ||::||||||
Db    449 SerIleAspAla-------------LeuLysPheGlyIleAspValLeu 460
```

```
Alignment Scores:

Pred. No.:                                          Length:    448
Score:              1e-15                            Matches:   125
Score:              309.50                           Conservative: 75
Percent Similarity: 42.64%                           Mismatches: 199
Best Local Similarity: 26.65%                        Indels:    70
Query Match:        11.21%                           Gaps:      20
DB:                 2
```

```
US-10-014-896-1 (1-1509) x A70578 (1-448)

Qy    150 GATGAAAGAGGCGGCTGAAAGGTGCCATCCAGATTCC------------------------ 185
          ||:||    |||    |||    ||||    |||    :|
Db      3 AspGluThrGlyAlaSerSerAspHisSerAspAspValAlaGlnValSerArgLeu 22

Qy    186 ------AACAGTGACTTTTAGCTTCTGAGAAGTCCAAATACTACAGCCCTGGCGTTGTT 236
          :::||||    |||    |||    |||    |||    |||
Db     23 IleArgPheAspThr-ThrAsnSerGlyLysProGlyThrGlyThrLysGlyThrProGluCy 42

Qy    237 CGGAAAATACATTCATAAAAGTGTTTTCCTACAGTGGTGTGTCCAGCACCAGCTTTATCCAGCATGA 296
          ::::||    |||    |||    |||    |||    :|    |||
Db     42 sAlaArgTrpValAlaGluGlnLeuAlaGluVal------------GlyTyrGlnProGl 58

Qy    297 AGTCGTGGGAAGAG----------TATAGGCCACCTGTTTCACT--ATCCAAGGCTCGGA 341
          |||||||    |||    :|    |||    |||    :||    |||
Db     58 uTyrValGluSerGlyValAlaProGlyAsnValPheAlaArgLeuAlaGlyValAlaAs 78

Qy    342 CCCCAACCTTGCAGCCTACTGTCACTTTGATGTGTGGTGCCTGCCCCGTGAAGA 401
          |||||||    |||    :|    |||    |||    |||
Db     78 pSerSerArgGlyAlaLeuLeuIleHisGlyHisLeuAspValValProAlaGluProAl 98

Qy    402 AGGCTGGGAGGTGCCCCCAATCTCTGGGTTGGAGGCTGATGCGGTCATCTATGGTGCGGGG 461
          |||    |||    |||    ||||    |||||||||||    :|
Db     98 aGluTrpTrpSerValHisProPheSerGlyAlaIleGluAspGlyTyrValTrpGlyArgGl 118
```

```
Qy    462 CACACTGGAGGACAAGAGAACTCTGATGATGGCCATTACTGCAGGTCCTTGGAGCCTCCTGCCTGAT 521
          ::::|||    |||    |||    |||    |||    |||||||||
Db    118 yAlaValAspMetLysSerAspMetValGlyMetMetIleValValAlaArgHisLeuArgGl 138

Qy    522 CAGGAAGTACATCCCCCGGAAGATCTTTCTTCATTTCTCTGGGCCATCATGATGAGGAGTCATC 581
          :::||||    |||    ||:    |||    |||    |||
Db    138 nAlaLeuAlaIleValProProArgAspProArgLeuValPheAlaPheValAlaAspGluHisGl 158

Qy    582 AGGGACAGGGGCTCAGAGGATCTCAGCCCTGCTCACAGTCAAGG---------------GG 626
          ||||    |||    |||    |||    :|    |||    :|
Db    158 yGlyLysTyrGlySerGlyHis-----TrpLeuValAspAsnArgProAspLeuPheAspGl 176

Qy    627 CGTCCAGCTAGCCCTTCATTGTGGACGGAGGGGGGGGGCTTCATCTTGGATGATTCATTCCTAA 686
          |||||||    |||    |||    |||||||||||||    |||
Db    176 yIleGluThrGlnAla---IleGlyGluValGlyGlyPheSerLeuThr------ValProAr 193

Qy    687 CTTC-------------AAGAAGCCCATGCCTTTGATTGCAGTCTCAGAGAAGAGGGTTCCATGAA 737
          |||    :::||||    |||    ||||||||||
Db    193 -------AAGAAGCCCATGCCTTTGATTGCAGTCTCAGAGAAGAGGGTTCCATGAA 213

Qy    738 CCTCATGCTGCAAGTAAAACAATGACTTCAGCCCACTCTTCAGCTCGCTCCAAAAGAGACAAG 797
          ||||||||||||||||||:|    |||    |||    |||
Db    213 pMetArgLeuThrAlaArgGlyArgAlaGlyLysGlySerMetValHisAspGlnAsnAl 233

Qy    798 CATTGGCATCCTTGCAGCTGCGTCAGCCGGATTGGGAGCAGACAACAATGCCTATACATATT 857
          |||||||||||:    |||    |||    |||    |||
Db    233 aValThrAlaValCysGluValAlaValProGlyAspGlyAspCysGlyAspAspAspValCy 253

Qy    858 TGGGAAGCGGGACAGTGGTGACTGTGATTGCAGCAACTGGGAAATGAGTTTCCCTTCCGT 917
          :::||    |||    |||    |||    :|
Db    253 s---ThrAspThrValAlaAspAspAspAspAspAspAspAspValAlaPh 272

Qy    918 CAAATATAAATCCTTAACCAATGCAATAAATATCAGGAACCAAACGCACTCACCATAATTCAAAGCAGG 977
          |||    |||    |||    :|    |||    |||    |||
Db    272 eAsp-------LeuAspSerProAspLeu-----AlaGlyThrIleAspGluGlyLysGluGlyProMe 289

Qy    978 AAATCCCTTAACCAATGCAATAAATAATCAGGACCACCAGCACTCACCATAATTCAAAGCAGG 1037
          |||    |||    |||    :::|||    ||||||||||
Db    289 tAlaArgMetLeuLysMetAlaValLeuHisAspThrAlaAsnProThrMetLeuLysAlaGl 309
```

```
Qy   1038 GGTCAAGTTCAAATGTCATCCCCCCAGTGGCCCACAGTCAACTTCCGGATTCACCC 1097
          :::||    |||    |||    |||||||||||
Db    309 yTyrLysValAlaAsnValValProAlaThrArgAlaGluValValAspCysArgValLeuPr 329

Qy   1098 TGGGACAG-------------ACAGTCCAAGAGGTCCTA-------GAAC/ACAGAA 1133
          ||||||    :|    |||    :|    |||    |||
Db    329 oGlyArgArgAlaAlaPheGluAlaGlnLeuValAspAlaLeuAlaLeuIleGlyProAspValThrAr 349

Qy   1134 GAACAATTGTGGCTGATAAACAGAGTCCAGTGCGTTCCATGTGGTCAGGTGGCCTTTGACCCCTCCC 1193
          |||    |||    :|    |||    :|    |||
Db    349 gGluTrpValSerAsp-------------------- 354

Qy   1194 CGTCAGCCCTTCTGATGACAGAGCGCCCTTGGGTGGTACCAGCTG------- 1244
          |||    |||    |||    :::|||    CTCCGCCGAC 1244
Db    355 ---LeuProSerTyrGluThrThrPheAspGlyAspLeuuValAlaLeuAlaMetAsnAlaAl 373

Qy   1245 CGTTCAGTCCCTTCCCTGGCGAAGTCAAATATTACTGCCCAGTTACTTCTTCTATTGGCAACAC 1304
          |||||||:||||||:    |||    |||    :|    |||
Db    373 aValLeuAlaValAspProAsp---GlyArgThrValProTyrMetLeuSerGlyGlyTh 392

Qy   1305 AGACAGCCGGATTCTTTACAAACCTCACCACTGGCACTCACGGTTGTCACCCCATCTACAT 1364
          ||||||||||    |||    :::|||    :|    |||
Db    392 rAspAlaLysAlaArgLeuGlyIleArgCysPheSerProLeuuArgLe 412

Qy   1365 ACAGGCCTGAA---GACTTCAAACGCATC---CATGGAGTCAACGAGGAAAATCAGTCCA 1418
          :::|||    ||:    |||    |||||||||||
Db    412 uProProGluAspProPheThrSerLeuPheHisGlyValAspGluArgValProIleAs 432

Qy   1419 AGCC-----TATGAGACCCAAGTG 1437
          :::    ||||||:||
Db    432 pGlyLeuArgPheGlyThrGluVal 440
```

T19182

hypothetical protein C10C5.4 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19182
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
A;Reference number: Z19085
A;Accession: T19182
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-397 <WIL>
A;Cross-references: EMBL:Z68214; PIDN:CAA92447.1; GSPDB:GN00022; CESP:C10C5.4
C;Genetics:
A;Experimental source: clone C10C5
A;Gene: CESP:C10C5.4
A;Map position: 4
A;Introns: 27/1; 48/3; 215/3; 268/1; 303/2; 330/3; 350/3; 383/1

Alignment Scores:

Pred. No.:	
Score:	1.41e-09
Percent Similarity:	229.50
Best Local Similarity:	38.18%
Query Match:	25.12%

Length:	397
Matches:	102
Conservative:	53
Mismatches:	184
Indels:	67
Gaps:	12

US-10-014-896-1 (1-1509) x T19182 (1-397)

Qy	286	ATCCAGCATGAAAGTGGTGAAGAG------------------------TATAGCCACCTGTTCACTATCCAA 333
Db	42	IleGluArgGlyAlaGlnIleGluThrArgAlaProGlyValTyrPheValIleMetThrIlePro 61

Qy	334	GGCTCGGACCCCAGCTTGCAGCCGTTACCAACTGCTGATAGGCTCACTTGATGGGTGCCTGCC 393
Db	62	GlySerGlnProGluLeuProSerIleMetLeuThrTyrSerHisThrAspValValProThr 81

Qy	394	CCTGAAGAAGGCTGGGGAGGTTGGCCCCCCATTCTTCGTGGGTTG---GAGCGTGAATGGGGGCTCATC 450
Db	82	PheArgGluHisTrpThrHisAspProTyrSerAlaPheLysAspGluAspGlyAsnIle 101

Qy	451	TATGGTTCGGGGCACACTGGACGACGAAGAGACTCTGTGATGGCATTACTCCGAGGCCGCCTTGGAG 510
Db	102	PheAlaAlaArgGlyAlaGlnAspThrLysCysLeuGlyValGlnHisIleValGlnIleAlaPheArg 121

Qy	511	CTCCTGCTGATCAGGAAGTACATCCCC---CGAAGACATCTTCTTCATTCATTCTCTGGGCCAT 567
Db	122	AsnLeuPheAlaLysGlyIleAlaLysGlnIleTrpLysArgThrIleHisIleValAlaValThrPrcGlyPro 141

Qy	568	GATGAGGAGTCATCAGGG---ACAGGGGGCTTCAGAGAGAGGTCTCAGCCCTGCTACAGTCAAGG 624
Db	142	AspGluGluThrGlyHisIleAsnGlyMetLysGlyPheValGluThrArgAsnGluPheLys 161

Qy	625	GGGGTCCAGCTAGCCTTCATTGTGACGAGGGGGCTTCATCTTGGATGAATTTCATTCCT 684
Db	162	LysLeuAsnIleGlyPheSerLeuAspGluGly---------His 173

Qy	685	AACTTCAAGAAGCCCATCCCCTTGATTGGCAGTCTCAGAGAGAGGTTCCATGAACCTCATG 744
Db	174	AlaSerGluAsnGlyValTyrLeuThrPheTyrAlaGluArgValThrThrpTrpPheLeuLys 193

Qy	745	CTGGCAAGTAAACATGACTTCAGGCCACTTCAGCTCCTCCAAAGGAGAGACAAGCCATTGGC 804
Db	194	ValThrArgAlaProGlyAspAsnProGlyHisGlyLySerGlnPheMetGluAsnThr------ 210

Qy	805	ATCCTTGCAGCTGTGTGTCAGCCACACCAAATGCCTATCATATATTTGGAAGC 864
Db	211	------AlaMetGlyAsnIleGluArg--------- 217

Qy	865	GGGACAGTGGTGACTTGACTTAGTGCAGCAACTGGCAAATGAGTTTCCCTGTCAAATATA 924
Db	217	-------------------- 217

US-10-014-896-1 (1-1509) x T19182 (1-397)

RESULT 9

AB1469

succinyldiaminopimelate desuccinylase homolog lin0289 [imported] - Listeria innocua (
C;Species: Listeria innocua
C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C;Accession: AB1469
R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloec
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi,
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maltournam, A.;
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehla
A;Title: Comparative genomics of Listeria species.
A;Reference number: AB1077; MUID:21537279; PMID:11679669
A;Accession: AB1469
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-378 <GLA>
A;Cross-references: GB:AL592022; PIDN:CAC95522.1; PID:g16412718; GSPDB:GN00178
A;Experimental source: strain Clip11262
C;Genetics:
A;Gene: lin0289

Alignment Scores:

Pred. No.:	
Score:	2.82e-09
Percent Similarity:	225.50
Best Local Similarity:	38.95%
Query Match:	23.23%

Length:	378
Matches:	102
Conservative:	69
Mismatches:	181
Indels:	87
Gaps:	16

Qy	925	ATCCTCGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATTGGAGAGAAATCCC 984
Db	218	PheLeuLeuAlaSerAlaAsnGluGlnLysGluLeuLeuGlyLysAsnPro 237

Qy	985	TTAACCAATGCAATAAATCAGGACCACCAGGCACTCACCAATATTCAAAGCAGGGGTCAAG 1044
Db	238	ThrTrpAlaLeuGlyAspValThrThrLeuAsnValAsnIleLeuLeuGlyValGln 257

Qy	1045	TTCAATGTCATCCCCCAGTGGCCCAGGCCACAGTCACTTCCGGATTCACCCTGGGACAG 1104
Db	258	PheAsnValIleProGlnGlyProGlnAlaThrValThrSerProAspSerProAsnGln 277

Qy	1105	ACAGTGCCAAGAGGTGCCTAGAAACTCACGAGAACAATTGGTGCTCAT------AACAGAGTC 1158
Db	278	AspPheGlyGluIleArgAlaLysLeuAspGlnTrpValLysAspAlaGlyValVal 297

Qy	1159	CAGTTCCATGTGTTGAGTGCCTTTGAACCCCCTCCCGTGCCAGCCCT-----TCTGATGAC 1212
Db	298	ThrTyrGluPheSerGlnHisSerAspTyrLysProValSerProHisThrArgAspAsp 317

Qy	1213	AAG-------GCCTTGGGCTACCAGCTGCTCCGGCCAGACCGTACAGTCCGGTCTTCCCGGAA 1266
Db	318	ProPheTrpAlaAlaPheGluAspSerLeuLysGlnLeuAsnCysGluPheThrThrGlu 337

Qy	1267	GTCAATATTACTGCCCCAGTTACTTCTTCATTGGCAACACAGACAGCGCGGATTCTTTACAAAC 1326
Db	338	Val-------------CysValGlySerThrAspSerArgPheValArgValArgLys 350

Qy	1327	CTCACCACTGCGCATCTACAGTTCTCACCCCATCTACACACTAGCCTGAAGACTTCAAACGC 1386
Db	351	AlaGlyValProAlaIleAsnPheSerProMetIleAlaAsnThrPro------ValArg 367

Qy	1387	ATCCAATGGAGTCAACGAGAGAAAAATCAGTCCAGCCTATGAAGACCCAAGTGAAATTCATC 1446
Db	368	AlaHisAlaHisAsnGluPheLeuAsnGlyLysValPheLeuGlyValGlyIlePhe 387

Qy	1447	TTTGAGTTGATTCAGAAT 1464
Db	388	GluThrLeuIleAsnAsn 393

US-10-014-896-1 (1-1509) x AB1469 (1-378)

RESULT 10

H90312

succinyl-diaminopimelate desuccinylase (dapE) [imported] - Sulfolobus solfataricus
C;Species: Sulfolobus solfataricus
C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
C;Accession: H90312
R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Ch
Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A;Molecule type: DNA
A;Residues: 1-410 <KUR>
A;Reference number: A99139
A;Accession: H90312
A;Status: preliminary
A;Description: Sulfolobus solfataricus complete genome.
A;Cross-references: GB;AE006641; NID:gi13814771; PIDN:AAK41759.1; GSPDB:GN00155
C;Genetics:
A;Gene: dapE

Alignment Scores:

Pred. No.:	
Score:	223.00
Percent Similarity:	40.43%
Best Local Similarity:	24.30%
Query Match:	8.07%
DB:	2

Length:	410
Matches:	113
Conservative:	75
Mismatches:	183
Indels:	94
Gaps:	22

US-10-014-896-1 (1-1509) x H90312 (1-410)

[The remainder of the page consists of a vertically-oriented nucleotide and amino acid sequence alignment between the query (Qy) and database (Db) sequences, with corresponding sequence position numbers.]

```
RESULT 11
C97868
hypothetical protein dapE [imported] - Rickettsia conorii (strain Malish 7)
C;Species: Rickettsia conorii
C;Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 22-Oct-2001

C;Accession: C97868
R;Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.;
Science 293, 2093-2098, 2001
A;Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A;Reference number: A97700; MUID: 21442074; PMID: 11557893
A;Accession: C97868
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-382 <KUB>
A;Cross-references: GB:AE006914; PIDN:AAL03885.1; PID:g15620491; GSPDB:GN00173
C;Genetics:
A;Gene: dapE
C;Superfamily: succinyl-diaminopimelate desuccinylase

Alignment Scores:
Pred. No.:
Score:                  214.50
Percent Similarity:     42.26%        Length:          382
Best Local Similarity:  24.08%        Matches:          98
Query Match:             7.77%        Conservative:     74
                                      Mismatches:      140
                                      Indels:           95
DB:                       2           Gaps:             23

US-10-014-896-1 (1-1509) x C97868 (1-382)

Qy   304  GAAGAGTATAGCCACCGTTCACTACCCCAAGGCTCG---GACCCCAGCTTGCAGCCCTAC  360
               ||:::  :::||||:||||::|||           |||  ::||||||||||||::||||
Db    49  GluGlnValThrAsnLeuTyrAlaValPheGlySerAsnGluProAsnIle-------    65

Qy   361  CTGCTCGATGGCTCACTTTGATGTGGTGCCTGCC---CCTGAAGAAGGCTGG---GAGGTG  414
               ||||||||||||||||||||||||||||||     |||||||||||||   |||||
Db    66  CysPheValGlyHisValAspValAlaLeuAlaGlyAsnHisGluLeuTrpHisAsnAla   85

Qy   415  CCCCCCATTCTCTGGGTTGGGAGGCGTGGGCGTCATCTATGGTCGGGGGCACACTGGACGAC  474
               |:::||||:||||:::||||||||||:||||||:::||||||||||||||||
Db    86  SerProPheLysValSerGlnGlnAspGlyLysIleLeuTyrGlyGlyArgGlyAlaValAspMet  105

Qy   475  AAGAACTCTGTGATGGGCATTACTGCAGGCC----TTGGACTTCGTGCTGATCGGAAGTAC  531
               ||:::||||||:::||||||||||        ||:::|||||||||||||||:||
Db   106  LysGlyLysAlaIleLeuAlaCysPheLeuAlaAlaSerLeuAspPheLysLysAsnThrAsp  125

Qy   532  ATCCCCGAGAGATCTTTCTTCATTTCTCTGGGCCATGAGGAGGAGTCATCATCAGGGACAGG  591
               |||||||||:|||:||:||||||||||||:::|||||||||||||:::||
Db   126  PheLysGlySerIleSerProPheLeuLeuThrSerAspGluGlyLysHisGlyGly        145

Qy   592  GCTCAGGAGAGATCTCAGCCCTTGCTACAGTCAAGGGACGGGGGCGGCGTCCAGCTAGCCTTC---ATTGTG  648
               |||:::||||||:::|||||||||||||:::|||||||||||||||||||||||||||||   |||:::
Db   146  ThrGlySerGluMetLeuGlnThrValArgIleTyrArgAspGlnGlyTyrLysIleAsnPheAlaLeuVal  165

Qy   649  GACGAGGGGGGCCTTCAYCTGGAYGAYGAATTCAYTCCTAACCTCAAGAAGCCCATCGCC---CCC  705
               |||||||||||||:::||||||:||:||:::||||:::|||||||||||||||||:::   |||
Db   166  GlyGlu------------------------------------------------------ProThrCysGlyLeu  176

Qy   706  TTGATTGCAGTCTCAGAAGAGGGGTTCCATGAAACCTCAYTGCTGCAAGTAAACATGACTTCA  765
               ||||:::||||||||:::||||||||||||:::||||:::||||||:::|||||:::||||:::
Db   177  AlaIleLysIleLeuTyrArgArgGlySerValAsnPheLysLeuAsnIleGluGlyLeuSer  196

Qy   766  GGGCCACTCTTCAGCTCCTCCCAAAGGAGAGACAAAGCATTGGCATCCTTGCAGCTGCTGTCAGC  825
               ||:::|||||||||||||||||||||:::|||||||||||:::||||||||||||||||:::
Db   197  GlyHisSerValAlaTyrProHisSerLysAlaAlaAsnAsn-----------------------  207

Qy   826  CGATTGGAGCGAGACCACCAATGCCTATCATATTTGGAAGCGGGGGACCAGTGGTGACTGTATTG  885
               ||||||:::|||||||||||||||::|||::||||||||||:::|||||||||||||:::
Db   208  -----------------------------------ProLeuProCysLeu------------    216

Qy   886  CAGCAACTGGCAAATGAGTTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTA  945
               |||||||||||||||:::||||||||||||:::|||||:::|||||||||||||
Db   217  AsnGluLeu--------------IleIleIleLeu                        227

Qy   946  TTTGAACACTTATAAGCAGGTTTATGGAGAGAAATCCCTTA------ACCAATGCAATA  999
               ||:::||||||||||||||||||||||:::|||||||||      :::|||||||
Db   228  -------------ThrGluPhePheGlnArgSerAsnLeuGluValThrPheAsnIleGlu  242

Qy  1000  ATCAGGACCACCAGGACCACTACCATATTCAAAGCAGGGGTCAAGTTCAAATCATCCCC  1059
```

```
Qy   508  GAGCTCCTGCTCGTGATCAGGAAGAAGTACATCCCCGAAGAGATCTTTCTTCATT---TCTCTGGGC  564
               ||| :::||| ::::|||
Db   137  GluMetLeuUArgArgAlaLysLeuLeuProSerAsnIleGlnValIleGlnThrIleVal  156

Qy   565  CATGATGAGGAGTCATCAGGAGACAGGGGCTCAGGAGGATCTCAGCCCTGCTACAGTCAAGG  624
               ::||||
Db   157  ProAspPGluGluThrValGly  165

Qy   625  GGCGTCCAGCTAGCCTTCATTCTTGGACGAGGGGGGCTTCATCTTGGATGAGATTCATTCCT  684
               :::|||||||||  ||||||||||||:::||
Db   166  AsnArg  165

Qy   685  AACTTCAAGAAGCCCATCGCC-----TTGATTGCAGTCTCAGAAGAGGGGTTCCATGAAC  738
               |||  :::|||||:::|||||||:::||
Db   185  IlePheThrGluProThrGlyProAspProAsnGlyHisArgGlyHisGlyAlaIleTrp  204

Qy   739  CTCCATGCTGCAAGTAAACATGACTTCAGGCCTCCTCCAAAGGAGAGACAAGC  798
               ||:::|||||:::||||||||||
Db   205  MetIleGluGlnLeuUTyrArgLysAlaIlePro------               220

Qy   799  ATTGGCATCCTTGCCTTCATTCTTGGACGAGGGGGGCAACCAAATGCCTATCATATTT  858
               :::||||:||||||||
Db   221  LeuGlyIleLeuAspLaVaLLysAlaLeuThrSerLeu                      231

Qy   859  GGAAGCGGGGACAGTGGTGACTGTATTGCAAATGAGTTTCCCTTCCCTGTC  918
               |||:::||||||:::|||||||||||||||||||
Db   232  MetLeuGluLeuUTyrLysAlaIleLeuPro                        241

Qy   919  AATATAATCCTGAGCAACCCATAGCTATTTGAACACTTATAAGCAGGTTTATGGAGA  978
               |||||||
Db   242  AspIleValSerLeuUTyr---GluIle                             249

Qy   979  AATCCCTTAACCAATGCAATAAATCAGGACCACCAGGACCACTATTCAAAGCAGG  1038
               |||  :::|||||||:||||||||||:::|||||:::|||||||||||||:::
Db   250  SerProGluSerGlyLys------LysProThrIleLeuUValGlyValLysCysGlyGly  267

Qy  1039  GTCAAGTTCAAATCATCCCCAGTGGCCCAGTGACACAACTTCCCGGATTCACCCT  1098
               |||||||:::|||||||:::|||||||||||:::||||||||||||||||:::
Db   268  ThrTrpValAsnThrValAlaAlaAspTyrCysGluPheTyrArgLeuUArgLeuUCys  287

Qy  1099  GGACAGCACAGTCCAAGAGGTC-----CTAGAACTCACGAGAAGAACATGTGGGCT  1146
               |||  :::|||  :::|||||||:::|||||||||||:::
Db   288  GluGluAsnLeuUGluGlnValLeuUVaLArgArgHisSerSerArgIleLeuIleArgValAlaLeuUArgValArgGlu  307

Qy  1147  GATACAGAGTCCAGTTCCATGTGTTGGTGAGTCCCTTTGACCCCGTCAGCCCTTCT  1206
               ||| |||||||||||||||||||:::
Db   308  SerThrGlyValArgTyrGlu---TyrAspGluPheTyrAlaVaLAsnThrMetArgCys  326

Qy  1207  GATGACAAGGCCTTGGGCTACCAGCGTCCGCCAGACCCGTACTCCGYCTTCCCGGAA  1266
               |||:::|||  ||||||||:::||||||
Db   327  AspAspGluUArgLeu---IleSerAlaLeuUArgGluLysIleGluVaL--------  342

Qy  1267  GTCAAATAATAACTGCC-----CCAGTTACTTCTTATTGGCAACAGAGACAGCCGAATTCTTTACA  1323
               |||:::||:::||||||||:::||||||||||||||||
Db   343  ArgGlyValGlyLuUAlaArgThrValLeuUSerAlaGlyThrPheAspIleArgPheThrArgIle  362

Qy  1324  AACCTCACCACTGGGCATCACTTACAGGTTCTACCATCACTACAGGGTTCAAGACTTCAA  1383
               ||| |||  ::::|||||:::|||||||||:::
Db   363  Asn------GluGlyIle-------LysSerIleAsnTyrGlyProGlyArggIleGlu  377

Qy  1384  CGCATCCATGGAGTCAACGAGAAAAATCTCAGTCCAAGCCTTATGACACCCAAGTGAAATTC  1443
               |||:::||||||:::||||||||||||||||||:::|||||||||||||||||:::
Db   378  LeuAlaHisAlaLAsnAspGluTyrIleTyrVaLLysAspLeuAspSerIleLysSerVal  397

Qy  1444  ATCTTTGAGTTGATT  1458
               |||:::||||:::
Db   398  LeuGlyLeuLeuVal  402
```

RESULT 12

E71074
hypothetical protein PH1288 - Pyrococcus horikoshii
C;Species: Pyrococcus horikoshii
C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C;Accession: E71074
R;Kawarabayasi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekir
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
DNA Res. 5, 55-76, 1998
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
A;Reference number: A71000; MUID:98344137; PMID:9679194
A;Accession: E71074
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-455 <KAW>
A;Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30391.1; PID:g3257708
A;Experimental source: strain OT3
A;Note: this accession replaces an interim accession for a sequence replaced by GenBank
C;Genetics:
A;Gene: PH1288
C;Superfamily: peptidase V

Alignment Scores:
Pred. No.:
Score: 2.51e-08
Percent Similarity: 213.50
Best Local Similarity: 38.91%
Query Match: 24.43%
DB: 7.73%
                          2

US-10-014-896-1 (1-1509) x E71074 (1-455)

Length: 455
Matches: 108
Conservative: 64
Mismatches: 145
Indels: 125
Gaps: 23

RESULT 13

C96507

hypothetical protein T12G22.9 [imported] - Arabidopsis thaliana (mouse-ear cress)

A;Species: Arabidopsis thaliana
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: C96507
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Raul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.;
Nature 408, 816-820, 2000
A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A;Reference number: A86141; MPID:21016719; PMID:11130712
A;Accession: C96507
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-438 <STO>
A;Cross-references: GB:AE005173; NID:g8655992; PIDN:AAF78265.1; GSPDB:GN000141
A;Gene: T12G22.9
A;Genetics:
A;Map position: 1

Alignment Scores:
Pred. No.:
Score: 3.24e-08
                                212.00
Percent Similarity: 37.04%
Percent Identity: 20.58%
Best Local Similarity: 7.68%
Query Match: 2
DB:

Length: 438
Matches: 100
Conservative: 80
Mismatches: 162
Indels: 144
Gaps: 19

US-10-014-896-1 (1-1509) x C96507 (1-438)

Qy   14 SerIleIlePheSerLeuGlnSerSerGluGluAspThrProIleThrArgPhe 33
         || ||| ||| ||||
Db 187 ACAGGTGACTTTTAGCTCTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGGACCCCAGCTG 237

Qy   34 GlnGlnThrTyrLeuArgPheAsnThrAlaHisProAsnProAsnTyrThrAlaProIleSer 53
         :::| |||
Db 250 --------CATAAAGTCTTTCCTACAGTTGGTCACGCACCAGCTTTATCCAG 291

Qy   54 PheLeuIleAsnGlnAlaGlnSerIleGlyLeuThrThrLysThrIleGluPheIleSer 73
         ||| ||||
Db 292 CATGAAGTCGTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGGACCCCAGCTG 351

Qy   74 GlyLysProIle 89
         :::
Db 352 CAGCCCTACCTGCTGATGGCTCACTTTGATGTGTTGGCCTGCCCCGTGAAGAAGGCTGGGAG 411

Qy   90 ProSerIleLeuPheAsnSerHisLeuAspSerValProAlaGluSerGlyLysTrpThr 109
         |||
Db 412 GTGCCCCAATTCTCTGGGTTGGAGCGT 468

Qy 110 TyrProProPheSerAlaHisLysThrIleAspGlyHisIleTyrAlaArgGlyAlaGln 129
         |||
Db 110 GATGGCGTCATCTAAGGTCGGGCACACTG 418

Qy 1309 AGCCGGATTCTTTACAAACCTCACCACTGGCGATCTACCCATCTACACACAG 1368
         |||||| ::|| :::|||||
Db 407 SerArgArgValTyrPheThr 416
         ProLeuGlyValLys 416

Qy 1369 CCTGAAGACTTC 1416
         -----AAAGGCATCATGGAGTCAACGAGAAAATCTCAGTC 1416
Db 417 AlaIleLeuAspIleGlyProArgGlyValGlyAsnIleHisGlyProAsnGluTyrValGluVal 436
         :::
         PheSerProLeuArgThrIleHisIleSerThrTyrValProGluGluGluIleGlyGlyPhe 436

Qy 1417 CAAGCC 1422
         :::
Db 437 AspSer 438

Qy 586 ACAGGGGCTCAGGAGGATCTCAGCCCTGCTACAGTCAAGGGCGGTCCAGGTTCATT 645
         ||| ::: ||| |||
Db 170 AspGlyMetMetLysPheAlaAlaSerSerGluPheLysAspLeuAsnLeuGlyPheAla 189

Qy 646 GTGGACGGAGGGG------------GGCTTCATCTTGGAGAGTTTCATTCCTAACTTCAG 693
         :::|||||||||| |||| :: ||
Db 190 MetAspGluGlyGlnAlaLysAsnProGly-------AspGluPhe 201

Qy 694 AAGCCCCATCGGCCTTGATTGCAGTCTCAGAGAAGAGGGTTCCAATGAACCTCATGGCCAAGTA 753
         ::| |||||| :::::
Db 202 --------ArgValPheTyrAlaAspArgValProTrpHisPheValIleIleLysAla 217

Qy 754 AACATGACTTCAGGCCCACTCTTCAGCTCCTCCAAAGGAGAGACAAGGCATTGGCATCCTTGCA 813
         ||| :::
Db 218 GluGlyTyrIleProGlyHisGlyAlaLys 226

Qy 814 GCTGCTGTCAGCCGATTGGAGCAGCAGACCAACAATGCCTATACTATTGGAAGGCGGACAGTG 873
         |||
Db 227 --------LeuTyrAspAsnSerAlaMet 233

Qy 874 GTGACTGTATTGCAGCAACTGGCAAAATGAGTTCTCCTTCCCGTCAATATAAATCCTGAGC 933
         :::::::::::   :::
Db 234 GluAsnLeuMetLysSerVal 240

Qy 934 AACCCATGGGCTATTTGAACCACTTATAAGGCAGGTTTATGGAGAGA---------- 978
         :::
Db 241 --------GluLeuIleSerArgPheArgGlnPheAspPheVal 254

Qy 979 -----------------AATCCTTA--ACAAATGCA 996
         |||||::: ||||| ::
Db 255 LysAlaGlyAlaGlyAlaTyrSerGluValIleSerArgValAlaProValTyrLeuLysAla 274

Qy 997 ATAATCAGGACCACCACGGGCACTCACCAATATTCAAAGCAGGGGTCAAGTTCAATGTCATC 1056
         ||||||||
Db 275 GlyThrProThrThrThr 287
         GlyPheValMetAsnMetGln 287

Qy 1057 CCCCCAGTGCGCCCAGGCCACAGTCAAACTTCCGGAATTCACCCT---------- 1098
         :::::::::   :::
Db 288 ProSerGluAlaGluAlaGlyTyrTyrAspLeuArgLeuProProMetAlaAspProVal 307

Qy 1099 ----------GGACAGCAGTCCAAGAGGTCCTAGAACTCACGAAGAGAACATT 1140
         :::   :::
Db 308 MetLysLysArgIleAlaGluTrpAlaProSerIleArgAsnMetThrTyrSerIle 327

Qy 1141 GTGGGTGATAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTTGACCCCGTCCCGTGAGCAGC 1200
         ||||||||||
Db 328 GlnGluLysGlyLeuArgAspHisLeu 343
         GlyArgProIle----MetThr 343

Qy 1201 CCTTCTGATGACAAGGCCTTGGGGTACAGCTGCTCCCAGACGGTACGGTCCGGCTTC 1260
         ||| |||
Db 344 ProValAsnAspSerAsnProTrpSerIlePheLysGlnAlaValGluAlaMetGly 363

Qy 1261 CCGGAAGTCAAATATTACTGCCCCAGTACTTCTATTGGCAACACAGACAGCAGCGGATTCTTT 1320
         ||||||||||
Db 364 GlyLys-----LeuAlaLysProGluIleLeuAlaSerThrThrAspSerArgArgPheIle 381

Qy 1321 ACAAAACCTCACCATCTACAGGTTTCTACCCCATCTACAACACCTGAAGACTTC 1380
         ||| |||
Db 382 ArgThrPheLeuGlyIleProThrPheGlyPheSerProMetThrAsnThrPro 398

Qy 1381 AAACGCCATCCCATTGGAGTCAACGAGAAAATC---------- 1419
         :::||| :::
Db 399 IleLeuLeuHisAspHisAsnGluPheLeuLysAspThrValPheMetLysGlyIleGlu 418

Qy 1420 GCCTATGAGACCCAAGTG 1437
         ||||||||||::||
Db 110 DATGGCGTCATCTAAGGTCGGGCACACTG 418

Qy 529 TACATCCCCGAAGATCTTCATTCTTCATTCTCTGGGGCATGTAGTGAGGAGCATCAGG--- 585
         ::: |||::: |||| ||||
Db 130 AspAspLysGlnArgValTyrLeuGlyLysValGlnThrTyrLeuGluSerIleArgAsnLeuLysSerArgArgGly 149

Qy 469 GACGACAAGAACTCTGTGATGGGCATTACTGCAGGCGCCTGGGAGGTCCTTGCTGATCAGGAAG 528
         || :::
Db 150 PheSerProLeuArgThrIleHisIleSerThrTyrValProGluGluGluIleGlyGlyPhe 169

```
Db         419 ValTyrGluSerValIle 424
                |||||| :::   ::: :
Qy             LeuTyrAspAsnSerAlaMetGluAsnLeuMetLysSerValGluLeuIleSerArgPhe 248

RESULT 14

A96506

probable aminoacylase [imported] - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C;Accession: A96506
R;Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
  J.; Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
R;Reference number: A86141; MUID:21016719; PMID:11130712
A;Accession: A96506
A;Status: Preliminary
A;Molecule type: DNA
A;Residues: 1-435 <STO>
A;Cross-references: GB:AE005173; NID:g11120814; PIDN:AAG30993.1; GSPDB:GN00141
A;Gene: T7O23.14
A;Map position: 1

Alignment Scores:
Pred. No.:              5.5e-08
Score:                  209.00
Percent Similarity:     38.96%
Best Local Similarity:  23.33%
Query Match:            7.57%
DB:                     2

US-10-014-896-1 (1-1509) x A96506 (1-435)

Qy   319 CTGTTCACTATCCAAGGCTCGGACCCCCAGCTTGCAGCCCTTACCTGCTGATGGCTCACTTT 378
         |||     |||             :::|||       ::|||             ||  :||:
Db    81 LeuLeuThrTrpLeuGlySerAsnLeuAsnLeuLysSerArgSerPheSerProLeuArgPheLeuAsnSerHisLeu 100

Qy   379 GATGTGGTGGTCGCCTGCCCCCTGAAGAAGGCTGGGGAGGTGCCCCCATTCTCTGGGGTTGGAGCGT 438
         |||| |||  |||       ||:|||||    :::  |||    ||        ||  |||||||||:
Db   101 AspSerValProAlaGluSerGluLysTrpIleHisProProPheSerAlaHisArgThr 120

Qy   439 ---GATGGCGTCATCTATATGTCGCGGGGCCACACTGGACGACAAGAACTCTGTGATGGCATTA 495
          :: ||||||||||     || ||||      |||           ||||        |||:
Db   121 IleAspGlyHisIleTyrAlaAargGlyAlaGlnAspAspSerCysIleGlyValGlnTyr 140

Qy   496 CTGCAGGCCCTGGAGCTCTCGATCAGGAAGTACATCCCCGGAAGATCTTCTTCATT 555
         |||||||||:|||:|||          ::: ||        ::|||
Db   141 LeuGlnAlaLeuGluLeuSerAsnLeuLysSerArgSerPheSerProLeuHisIleIleLeuPheAsnSerHisIle 160

Qy   556 TCTCTGGGCCATGATGAGGAGGTCATCAAGGAGGACAGGGGCT---CAGAGGATCTCAGCCCTG 612
         :: |||||||       |||||||||||||||||||||||||    |||||||||||||
Db   161 SerTyrValProGluGluGluIleGlyMetMetLysPheAlaAlaSer 180

Qy   613 CTACAGTCAAGGGGCGTCCAGCCTTCAATTGTGGACGAGGGG------------GGC 660
         |:::             |||:||||||||||||||||                |||
Db   181 SerGluPheLysAspLeuAsnLeuGlyPheValMetAspGluGlyGlnAlaLeuSerProGly 200

Qy   661 TTCACTCTTGGATGATTTCATTCCTAACTTCAAGAAGACCCATCGCCTTGAATTGCAGTCTCA 720
         |||             :: |||                        ||:       |||
Db   201 ------AspGluPhe 208

Qy   721 GAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCT 780
         |||||||        |||            :::                  |||
Db   209 GluArgThrProTrpHisLeuAlaIleLeuArgAlaAsnGlyMetProGlyHisGlyAlaLys 228

Qy   781 CCTCCAAAGAGAGCAACAAGCAATTGGCACCCTTCCAGCTGTC--------AGCCGATTG 831
         :::||||||::::     ||:::     ::|||||      |                  ||||||
```

```
Db         229 LeuTyrAspAsnSerAlaMetGluAsnLeuMetLysSerValGluLeuIleSerArgPhe 248
                   ::: |:   :
Qy         832 GAGCAGCAGCACCAATGCCTATCATCATATATTTGGAAGCGGGACAGTGGTTGGACTGTATTGCAGCAA 891
                                                                               ::: --Lys 260
Db         249 ArgGluSerGlnPheAspPheValLysAlaGly- 260

Qy         892 CTGGCAAAATGAGTTTCCTTCCCTCGTCAATATAATCCTAGCAACCCATGGCTATTTGA 951
                :: : ||                                   :::
Db         261 AlaAlaAsnSer--------------------------------GluValIleSerValAsnPro-- 271

Qy         952 CCCACTTATAAGCACAGGTTTATGGAGAGAGAAATCCCTTAACCAATGCAATAATCAGGAACACC 1011
                                                            :::::
Db         272 -------------------------------ValTyrIleLeuTyrAlaGly 277

Qy        1012 ACGGCACCTCACCATATTCAAAGCACAGGGGTCAAGTTCATCCCCAGTGGCCCAG 1071
                :::
Db         278 ThrProSerThr----------ThrGlyPheValMetAsnMetGlnProSerGluAlaGlu 294

Qy        1072 GCCACAGTCAACTTCCGGATTCAACCCT---- 1098
                |||:||  |||
Db         295 ValGlyTyrAspLeuArgLeuProProMetAlaAspProValAlaLeuAlaLeuLysSerArgIle 314

Qy        1099 ---GGACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACAATTGTGGCTGATAAACAGA 1155
                :::                                            :: ||| |||
Db         315 AlaGluGluThrpAlaProSerIleArgAsnMetThrTyrThrLeuLysSerGlnLysGlnLys 334

Qy        1156 GTCCAGTTCCATGTGTTGAGTGTGCCTTTGACCCCCTCCCAGCCGTCAGCCGTTCTGATGAAGCAAG 1215
                :::                                       ::|||          || |||
Db         335 LeuThrAspHisLeu--------GlyArgProIleMetGluMetThrThrThrAsnAsp-ThrAs 351

RESULT 15

T19180

hypothetical protein C10C5.3 - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T19180
R;Matthews, P.
submitted to the EMBL Data Library, December 1995
R;Reference number: Z19085
A;Accession: T19180
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-399 <WIL>
A;Cross-references: EMBL:Z68214; PIDN:CAA92445.1; GSPDB:GN000022; CESP:C10C5.3
A;Experimental source: clone C10C5
C;Genetics:
A;Gene: CESP:C10C5.3
A;Map position: 4
A;Introns: 27/1; 48/3; 215/3; 268/1; 303/3; 330/2; 350/3; 383/1

Alignment Scores:
Pred. No.:        6.99e-08
Score:            207.50
```

```
Qy        1216 GCCTTGGGCTACCAGCTGCTCGGCCAGAGACGGCTACAGCTCCGGTCCTCCGGAAGTCAATATT 1275
                ||||||         :: :||              ::|||         ||:
Db         335 ----------------GlyArgProIleMetGluMetThrThrThrAsnAsp-ThrAs 351

Qy        1276 ACTGCCCCAGTTACTTCTATTGGCAACACACACAGCAGCGAGAGCGGATTCTTTACAAACCTCACCACT 1335
                |||                  ::                   |||||||||||||||||
Db         351 nProTrp--TrpSerIlePheLysGlnAlaLeuValGluAlaThrGlyGlyLys----- 368

Qy        1336 GGCATCTACAGGTTGTTACCCCATCTACATACACCAGCCTGAAGACTTCAAAGCCATCATTGGA 1395
                :::                                                            :::
Db         369 AlaLysProGluIleLeuIleLeuSerThrThrAspSerArgPheIleLeuGluLysArgThrLeuLeuGlyIle 388

Qy        1396 GTCAACGAGAAAATC----------TCAGTCCAAGCCTATGAGACCCAA 1434
                |||||||                  ||||| ||||||||||
Db         389 ProValLeuGlyPheSerProMetIleLeuAsnThrPro--------IleLeuValHisAsp 405

Qy        1435 GTG 1437
                :::
Db         426 Ile 426
```

```
Percent Similarity:    38.65%     Conservative:   68
Best Local Similarity: 23.80%     Mismatches:    188
Query Match:            7.51%     Indels:         93
DB:                         2     Gaps:           18

US-10-014-896-1 (1-1509) x T19180 (1-399)

Qy    136 GAGGAAGACGGTGCGGATG---------------AAAGAGGGGCTGAAAGTGCCATCCAGATTCCA 186
              ||||||||||||||||||                 ::|  ||||||| |||||| ||||||||||
Db      3 GluGlnHisIleLeuAlaValThrArgPheArgGluThrTyrLeuArgValAsnThrGluGlnPro 22

Qy    187 ACAGTGACTTTTAGCTTCTGAGAAGTCAAATACTACTAGCCCTGGCTGAGTTCGGAAAATAC 246
              ::||||||||||||                 |||||:
Db     23 AsnProAspTyrAlaAlaLysArgAspPheLeuPheLysTyrAlaAspGluLeuGlyIle 42

Qy    247 ATTCATAAAGTCTTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGAGTCGTGGAA 306
              ||||  ||| ||||||||              |||:
Db     43 AlaArgArgSerPheGluThrValProGlyAlaIlePheVal-------------- 56

Qy    307 GAGTATAGCCACCTGTTCACTATCCAAAGGCTCGGACCCCACCCTTGCAGCCCTACCTGCTG 366
              |||||||||||||||||              ||| |||||| ||||||||||||
Db     57 ------------IleMetThrIleProGlySerGlnProGlyLeuProSerIleMetLeu 72

Qy    367 ATGGCTCACTTTGATGTGGCCTGCCCTGAGAAGGCTGGGAGGTGGCCCCATTCTCT 426
              |||||||||||||||                 ||||||||||||||||||:|||
Db     73 TyrSerHisThrAspValValProThrProPheArgGluHisTrpThrHisAspProTyrSer 92

Qy    427 GGGTTG----GAGCGTGATGGCCGTCATCTATGGTCGGGGGCACACTGGACGACAAGAGACTCT 483
              :: ||      ||||||| |:||:   |||| ||||||||        ||| |||
Db     93 AlaPheLysAspGluAspGlyAsnIlePheAlaAlaArgGlyAlaGlnAspMetLysCysVal 112

Qy    484 GTTGATGCATTACTGCAGGCCTTGGAGCTCCTGCTG--------ATCAGGAAGTACATC 534
              |||| |||||||||              ||||||||         |||||||||||
Db    113 GlyValGlnIleMetGluAlaLeuArgAsnLeuPheAlaGlnGlyIleArgGlnTrp--- 131

Qy    535 CCCGAAGATCTTCTTCAATTTCTGGGCCCATGAGAGAGGAGTCATCAGGG---ACAGGG 591
              ::||||||||:|||||              ||| |||:|||||||||||   ||||||
Db    132 ---LysArgThrIleHisLeuValTrpGlyProAspGluIleIleGlyIleAsnGly 150

Qy    592 GCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGCGTCCAGCTACCTTCATTGGAC 651
              ||||||||||                 ::| |||||||||||||||||||||||||||
Db    151 MetLysGlyPheAlaLysThrAspGluPheLysLysAsnLeuGlyPheSerLeuAsp 170

Qy    652 GAGGGGGGCTTCATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGCCTTGATT 711
              :::|||||||||                 :||    :||||||||||||:||
Db    171 GluGlyMetProSerAspSpAsp------------ValTyrLysVal 182

Qy    712 GCAGTCTCAGAGAAGGGGTTCCATGAACCTCATGCTGCAAGTAAAACATGACTTCAGGCCAC 771
              :::||||||| |||              :|||     ||| ||||||
Db    183 PheTyrAlaGluArgValAlaLaTrpIleProGlyAsnProGlyHis 202

Qy    772 TCTTCAGCTCCTCCAAAGGACAACAAGCCATTGGCATCCTTGCGCTGTCAGCCGGATTG 831
              |||||                 ::: |||||||:|||:::||
Db    203 GlySerGlnPheMetGluAsnThr------------AlaMetGluLysLeu 215

Qy    832 GAGCAGACACCAAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAA 891
              ||||||||||||||||
Db    216 GluArg------------------ 217

Qy    892 CTGGCAAAATGAGTTCCCTTCCCTGTCAAATAAATCCTGAGCAACCCATGGCTATTTGAA 951
              |||||||||||              |||  ||:||   :|||||||::||||||||||
Db    218 -----------TyrLeuAlaSerAlaArgGlysPheArg 226

Qy    952 CCACTTATAACCAGGTTATGTGAGAGAAAATCCC--TTAACCAATGCAATAAATCAGGACC 1008
              :::||||||              ||||| ||||||   :|| |||||||||||
Db    227 AspGluGlnIleLysAlaLeuLeuGluSerAsnProAspLeuThrIleGlyAspVal--- 245

Qy   1009 ACCACGGCCACTCACCATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCC 1068
              |||                 |||
Db    246 ThrLeuAsnValAsnIleValAsnGlyTyrValGlnPheAsnValIleProGlyLysPhe 265

Qy   1069 CAGGCCACAGTCAACTTCCGGACAGACAGTCCAAGAGAGTCCTAGAAACTC 1128
              ::||||              |||              |||
Db    266 GluAlaPheValAspIleArgLeuThrProSerIleAspPheAsnGluMetArgAsnLys 285
```

```
Qy   1129 ACGAAGAACATTGTGGCTGAT--------------AACAGAGGTCCAGTTCCATGTGTTGAGGTGCCTTTT 1182
              |||                                   :::||||:||||||||||||||||:|||
Db    286 LeuAspGlnTrpValLysAspAlaGlyLysGluGlyLeuValThrTyrGluPheSerLysLysHisSer 305

Qy   1183 GACCCCTCCCCGTCAGCCCT---------------TCTGATGACAAGGCCTTGGGCTACCAG--- 1230
              |||                                 ||:||||||||||||||||||||||
Db    306 AspLeuLysLeuSerValThrProHisThrArgAspAspProPheTrpValAlaPheGluAsp 325

Qy   1231 CTGCTCCGCCAGACCGTACAGTCCGTCTTCCCGGAGGTCAATAATTACTGCCCAGTTACT 1290
              |||:::||||||              |||              |||
Db    326 SerLeuLysGlnGluGlyCysLysSerPheThrThrGluVal------------- 338

Qy   1291 TCTTATTGGCACACAGCACACACAGCCGATTCTTACAAACCTCACCACGTGGCATCTACAGGTC 1350
              |||                 |||
Db    339 LeuIleGlySerThrAspSerArgIleValArgGlyValArgGlyAlaIleAlaAsnPhe 358

Qy   1351 TACCCCATTCACATACAGCGCCT-----------GAAGACTTCAAAACGCATCCAT 1392
              ||||||||||||||||||:::|||           :::|||
Db    359 SerProLeuIleAsnThrProLeuLeuAlaHisIleAlaHisAsnGluPhe----- 374

Qy   1393 GGAGTCAACGAGAAAATC----------TCAGTCCAAGCCTTATGAGACCCCAAGTG 1437
              ::|||||||||||:::                |||:::
Db    375 ---LeuAsnGlyLysValPheLeuAsnArgGlyIleGlnIlePheTyrGlnThrLeuIle 391
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: June 27, 2003, 10:51:58 ; Search time 28 Seconds
(without alignments)
4470.556 Million cell updates/sec

Title: us-10-014-896-1
Perfect score: 2762
Sequence: 1 atggctaagcggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table: BLOSUM62
Xgapop 10.0 ; Xgapext 0.5
Ygapop 10.0 ; Ygapext 0.5
Fgapop 6.0 ; Fgapext 7.0
Delop 6.0 ; Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104413_10368/app_query.fasta_1.1671
-DB=SwissProt_40 -QFMT=fastan -SUFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10014896_@CGN_1.1_34_@runat_27062003_104413_10368 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -LONGLOG
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	540	19.6	576	1	CBPS_YEAST	P27614 saccharomyc
2	202	7.3	406	1	ACY1_PIG	P37111 sus scrofa
3	199	7.2	410	1	Y457_METJA	Q57899 methanococc
4	185.5	6.7	408	1	ACY1_HUMAN	Q03154 homo sapien
5	180	6.5	375	1	DAPE_BUCAI	P57196 buchnera ap
6	170	6.2	377	1	DAPE_HAEIN	P44514 haemophilus
7	163.5	5.9	375	1	DAPE_ECOLI	P24176 escherichia
8	160	5.8	878	1	YB9X_YEAST	P38149 saccharomyc
9	149	5.4	470	1	PEPV_LACDL	P45494 lactobacill
10	146.5	5.3	403	1	YGEY_ECOLI	P46805 escherichia
11	143.5	5.2	381	1	ARGE_BUCAI	P57155 buchnera ap
12	138.5	5.0	383	1	ARGE_ECOLI	P23908 escherichia
13	134.5	4.9	481	1	YFL4_YEAST	P43616 caenorhabdi
14	116	4.2	3178	1	YS89_CAEEL	Q09624 caenorhabdi
15	115.5	4.2	383	1	HIPO_CAMJE	P45493 campylobact
16	114	4.1	890	1	AYS8_HUMAN	Q9up79 homo sapien
17	113	4.1	421	1	Y236_METJA	Q57688 methanococc
18	108	3.9	1231	1	BLM_CAEEL	O18017 caenorhabdi
19	107	3.9	1305	1	RPOC_UREPA	Q9pqv5 ureaplasma
20	106.5	3.9	1140	1	YM96_YEAST	Q04893 saccharomyc
21	106	3.8	551	1	YG1F_YEAST	P53214 saccharomyc
22	105.5	3.8	454	1	PUCC_RHOSU	P95556 rhodovulum
23	105	3.8	552	1	ASO_CUCPM	P37064 cucurbita p
24	104.5	3.8	643	1	PGT_RAT	Q00910 rattus norv
25	104.5	3.8	1306	1	MSB2_YEAST	P32334 saccharomyc
26	104.5	3.8	2472	1	NCR2_MOUSE	Q9wd42 mus musculu
27	104	3.8	1502	1	N170_YEAST	P38181 saccharomyc
28	103.5	3.7	892	1	DRN2_PIG	O62855 sus scrofa
29	103.5	3.7	364	1	ATX7_HUMAN	O15265 homo sapien
30	103	3.7	2845	1	APC_MOUSE	Q61315 mus musculu
31	103	3.7	977	1	YAG9_SCHPO	O09872 schizosacch
32	102.5	3.7	369	1	DAPE_CORGL	Q59284 corynebacte
33	102.5	3.7	1001	1	RR44_YEAST	O00162 saccharomyc
34	102.5	3.7	1429	1	EXPA_DROME	Q07436 drosophila
35	102	3.7	3969	1	HRX_HUMAN	Q03164 homo sapien
36	101.5	3.7	499	1	C719_SOYBN	O81970 glycine max
37	101.5	3.7	691	1	PP2L_YEAST	P26570 saccharomyc
38	101	3.7	571	1	ILVX_BACSU	Q04789 bacillus su
39	101	3.8	1367	1	AMYH_YEAST	P08640 saccharomyc
40	100.5	3.6	1377	1	NTRB_BRASR	P10578 bradyrhizob
41	100.5	3.6	861	1	IMB1_YEAST	Q06142 saccharomyc
42	100	3.6	1609	1	FIG2_YEAST	P25553 saccharomyc
43	100	3.6	462	1	TUBE_DROME	P22812 drosophila
44	100	3.6	3381	1	PGCV_BOVIN	P81282 bos taurus
45	99.5	3.6	579	1	ASO_CUCMA	P24792 cucurbita m

ALIGNMENTS

RESULT 1
CBPS_YEAST

ID   CBPS_YEAST   STANDARD;   PRT;   576 AA.
AC   P27614;
DT   01-AUG-1992 (Rel. 23, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Carboxypeptidase S precursor (EC 3.4.17.4) (YSCS) (GLY-X
DE   carboxypeptidase).
GN   CPS1 OR CPS OR YJL172W OR J0510.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AB320;
RX   MEDLINE=91243870; PubMed=1709881;
RA   Bordallo J., Bordallo C., Gascon S., Suarez-Rendueles P.;
RT   "Molecular cloning and sequencing of genomic DNA encoding yeast
RT   vacuolar carboxypeptidase yscS.";
RL   FEBS Lett. 283:27-32(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=91224132; PubMed=2026161;
RA   Spormann D.O., Helm J., Wolf D.H.;
RT   "Carboxypeptidase yscS: gene structure and function of the vacuolar
RT   enzyme.";
RL   Eur. J. Biochem. 197:399-405(1991).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Obermaier B., Piravandi E., Rinke M., Domdey H.;
RL   Submitted (SEP-1995) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SUBCELLULAR LOCATION.
RX   MEDLINE=92235008; PubMed=1569061;
RA   Spormann D.O., Helm J., Wolf D.H.;
RT   "Biogenesis of the yeast vacuole (lysosome). The precursor forms of
RT   the soluble hydrolase carboxypeptidase yscS are associated with the
RT   vacuolar membrane.";
RL   J. Biol. Chem. 267:8021-8029(1992).

RN   [5]
RP   SIMILARITY TO ARGE/DAPE/ACY1/CPG2/YSCS FAMILY.
RX   MEDLINE=94245187; PubMed=8188249;
RA   Henikoff S., Henikoff J.G.;
RT   "Protein family classification based on searching a database of
RT   blocks.";
RL   Genomics 19:97-107(1994).
CC   -!- FUNCTION: NECESSARY FOR USE OF CERTAIN PEPTIDES AS SOLE NITROGEN
CC   SOURCE. MAY ALSO CLEAVE INTRACELLULARLY GENERATED PEPTIDES TO
CC   RECYCLE AMINO ACIDS FOR PROTEIN SYNTHESIS.
CC   -!- CATALYTIC ACTIVITY: Peptidylglycine + H(2)O = peptide + glycine.
CC   -!- PATHWAY: NITROGEN METABOLISM.
CC   -!- SUBUNIT: YSCS IS SYNTHESIZED AS ONE POLYPEPTIDE CHAIN PRECURSOR
CC   WHICH AFTER CARBOHYDRATE MODIFICATION IN THE SECRETORY PATHWAY
CC   YIELDS TWO ACTIVE PRECURSOR MOLECULES. THE PROTEOLYTICALLY
CC   UNPROCESSED FORMS ARE ASSOCIATED WITH THE MEMBRANE, WHEREAS THE
CC   MATURE FORMS OF THE ENZYME ARE SOLUBLE.
CC   -!- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES.
CC   -!- PTM: GLYCOSYLATED.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X57316; CAA40571.1; -.
DR   EMBL; X63068; CAA44790.1; -.
DR   EMBL; Z49447; CAA89467.1; -.
DR   PIR; S16881; S16881.
DR   PIR; S16693; S16693.
DR   MEROPS; M20.002; -.
DR   SGD; S0003708; CPS1.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Hydrolase; Carboxypeptidase; Glycoprotein; Signal.
FT   SIGNAL   1   39   POTENTIAL.
FT   CHAIN   40   576   CARBOXYPEPTIDASE S.
FT   CARBOHYD   88   88   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   176   176   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   228   228   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   381   381   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   525   525   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD   387   387   N-LINKED (GLCNAC...) (POTENTIAL).
FT   CONFLICT   387   387   S -> T (IN REF. 2).
SQ   SEQUENCE   576 AA; 64597 MW; 5CBB536D42185F70 CRC64;

US-10-014-896-1 (1-1509) x CBPS_YEAST (1-576)

Alignment Scores:
Pred. No.:
Score: 1.92e-32
Score: 540.00
Percent Similarity: 47.49%
Best Local Similarity: 31.86%
Query Match: 19.55%
DB: 1

                        Length: 576
                        Matches: 159
                        Conservative: 78
                        Mismatches: 184
                        Indels: 78
                        Gaps: 19

Qy   157 GAGGCGCTTGGAAAGGTGCCATCCAGATTCCAACAGTGACTTTAGCTCTGGAAGAGTCCAAT 216
         ||||||| || ||||| || :|||||||||||
Db   92 GluLysLeuSerAsnAlaValArgIleProThrValVal-----GlnAspLysAsnPro 109

Qy   217 ACTACAGCC----------------------------------CTGGCTGAGTTCGGAAAATACATTCAT 252
         ||| ||| |||
Db   110 AsnProAlaAspAsp--ProAspPheTyrLysHisPheTyrGluLeuHisIleAspTyrPheGlu 129

Qy   253 AAAGTCTTTCCTACAGTGGTGGTCGGACCACCAGCTTATCCAGCAATGAAGAGTCGTGGAAGAGTAT 312
         ||| |||||||| |||||||| ||||| ||||| || :|| |||
Db   130 LysThrPheProAsnIle-----------HisLysHisLeuLeuLysValAsnGluLeu 147

Qy   313 AGCCACCTGTTCACATATCCAAGGCTCGGACCCCAGCTTGCAGCCCTTACCTTGCTGATGGCT 372
         ||| :|:||| :|| ||||| :||||| |||||||||||||| ||||
Db   148 GlyLeuLeuTyrThrTrpGluGlySerAspProAspLeuLeuLeuLeuMetAla 167

Qy   373 CACTTTGATGTGGTGGTGCCTGCCTGCCCTGAAGAT-------GGCTGGGAGGTGGCCCCCATTC 423
         :|||||||||||||||||| |||||| |||::: |||||||||||||
Db   168 HisGlnAspValValProCysProGluAspAspLeuThrLeuSerSerTrpLysPheProProPhe 187

Qy   424 TCTGGG---------------TTGGAGCGTGATGGGCGTCATCTANGGTCGGGGGCACACTGGACGAC 474
         ||| ||| ||| :|| |||
Db   188 SerGlyHisThrTyrAspProGluThrAsp---PheValTrpLysThrGlySerAsnAspCys 206

Qy   475 AAGAACTCTGTGATGGCATTACTGGCAGGCCGTTGGAGCGTCCTGCTGATCAGGAAGTACATC 534
         |||||| :|||||||||: ||||| :|||||||||||||| :||||||||
Db   207 LysAsnLeuLeuIleAlaGlyAspPheGlnAlaIleGluGlnLeuLeuIleAspGlyPheLys 226

Qy   535 CCCGGAAGATCTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCAGGGACA----GGG 591
         |||||| :||:::|||||| ||||||||||||||:||||||| |||
Db   227 ProAsnArgThrIleValMetSerLeuGlyPheAspGlyGluAlaSerGlyThrLeuGly 246

Qy   592 GCTCAGAGGANTCTACAGGCCCCTTGCTACAGTCAAGG----------GGGCGTCCAGCTAGCC 639
         ||| ||||||| |||||||||| ||||||||||
Db   247 AlaLeuAlaSerLeuAlaSerPheLeuLeuHisGlyArgThrTyrGlyTyr------ 264

Qy   640 TTCATTGTGGACGAGGAGGGGGGGGCTTCAATCTTG-----------GATGATTTCAATTCCTAC 687
         |||||| ||| :|||||| ||| :::||||||
Db   265 SerIleIleAspGluGlyGlyIleLeuMetGluValAlaAspLysAspValPheValAlaThr 284

Qy   688 TTCAGAAGGCCCATCCCGTTGAATCGAGTCTCAGAAGAGGGGTTCCATGAACCTCATGCTG 747
         ||| :::|| ::|||||||||||||||:|||||::::|||||||
Db   285 --------ProIleAsn-------AlaGluLysGlyTyrValAspPheGluVal 297

Qy   748 CAAGTAAACATGACTTCAGGCCCACTCTTCAGCTCTCCAAAGAGAGCAAAGCAATTGGACATC 807
         ||| |||||| ||||| |||||||:|||||||||:|||||
Db   298 SerIleLeuGlyHisGlyHisSerSerValProProAspHisThrThrIleGlyIle 317

Qy   808 CTTGCAGCTGCTGTCAGCCGATTGGAGCAGCAGAACCACCAATGCCATCAATATTTGGAAGCGGG 867
         ::: ||| ||| |||::: |||||||
Db   318 AlaSerGluLeuIleThrGluPheCysProAlaProArgLysAspLysAspPheAsp 337

Qy   868 ACAGTGGTGGACTGTATTGCAGCAACTGGCAAAGTGAGTTTCCCTGTGTCAATATAATC 927
         ||| :::|| :::||||||||||||||||||:::||||||::::|||||||
Db   338 AsnProIleTyrGlyIleuLeuThrCysAlaAlaGluHisSerLeuSerLeuLysAsp 357

Qy   928 CTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGG----966
         ||||||| ||| ||| |||
Db   358 ValLysLysThrIleLeuGlyAlaProPheCysProArgLysAspLysAsnLeuValGlu 377

Qy   967 TTTATGGAGGAGAAAATCCCTTAACCAAATGCAATAANTCAGGAGCAGCCACCAGGGCACTCACCATA 1026
         ||| ||||||| ||||||||||:::
Db   378 TyrIleSerAsnGlnSerHisLeuArgThrThrGlnAlaValAspIle 397

Qy 1027 TTCAAAGCAGGGGTCAGGTTCAATGTCATCCCCCAGTGGCCCAGGCCAGTCACTTC 1086
         |||||||| :::|| |||
Db   398 IleAsnGlyValLysAlaAsnAlaLeuProGluThrThrArgPheLeuIleAsnHis 417

Qy 1087 CGGATTCACCCTGGACAGACAGTCCAAGAGTGCCAGAA----1125
         ||||||||| :::||| |||
Db   418 ArgIleAsnLeuHisSerSerValAlaGluValPheGluArgAsnIleGluTyrAlaLys 437

Qy 1126 --------------------------------CTCACGAAGAAC------ATTGTGGCT 1146
                                         |||||||||| :::|||
Db   438 LysIleAlaGluLysTyrGlyTyrGlyIleuSerLysAsnGlyAspTyrIleIlePro 457

Qy 1147 GAT-------------AACACGAAGTTCAGTGTCCAATGTGTGAGTGCCTTGACCCCGTC 1197
         ::: ||| |||
Db   458 GluThrGluLeuGlyHisIleAspIleThrLeuLeuArgGluLeuProLeuProLeu 477

Qy 1198 AGCCCTTCTGATGACAGACAGAAGGCCTTGGGCTCCGGCCGAGACGTCCAGTCCGTC 1257
         ||||||||| || :::||| :::|| :::|||
Db   478 SerProSerSerGlyProVal------TrpAspIleLeuLeuAlaGlyThrIleGlnAspVal 495

```
Qy    1258 TTC---------------------------CCGGAAGTCAATTACTGCCCCAGTTACTTCT 1293
           |||                           :::||:||                  |||
Db     496 PheGluAsnGlyValLeuGlnAsnAsnGluGluPheTyrValThrArgGlyLeuPheSer 515

Qy    1294 ATTGGCAACACAGACAGCCGAATTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTAC 1353
           ||||||||||||||||||||||||:::||||||||||||||||||||:||
Db     516 ---GlyAsnThrAspThrLysTyrGlyTyrThrProAsnLeuSerLysAsnIleTyrArgPheVal 534

Qy    1354 CCCATCTACATACAGCCCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAGAAATCTCA 1413
           |||                ||                ||||||||   ||||||||||
Db     535 GlySerIleIleAspIleAspLeuLysThrLeuHisSerValAsnGluHisValAsp 554

Qy    1414 GTCCAAGCCTATGAGAGCCCAAGTGAAATTCATCTTTGAGTTGATTGAATGCTGAC 1470
           ||||:::   ||||||||:||:||||||      |||||||   ||| ||
Db     555 ValProGlyHisLeuSerAlaIleAlaPheValTyrGluTyrIleValAsnValAsn 573
```

RESULT 2

```
ACY1_PIG
ID    ACY1_PIG     STANDARD;     PRT;   406 AA.
AC    P37111;
DT    01-OCT-1994 (Rel. 30, Created)
DT    01-OCT-1994 (Rel. 30, Last sequence update)
DT    15-JUN-2002 (Rel. 41, Last annotation update)
DE    Aminoacylase-1 (EC 3.5.1.14) (N-acyl-L-amino-acid amidohydrolase)
DE    (ACY-1).
GN    ACY1.
OS    Sus scrofa (Pig).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX    NCBI_TaxID=9823;
RN    [1]
RP    SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC    TISSUE=Kidney;
RX    MEDLINE=93183401; PubMed=1292507;
RA    Jakob M., Miller Y.E., Roehm K.H.;
RT    "Cloning and sequence analyses of cDNAs encoding aminoacylase I from
RT    porcine kidney.";
RL    Biol. Chem. Hoppe-Seyler 373:1227-1231(1992).
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=93203140; PubMed=1284246;
RA    Mitta M., Ohnogi H., Yamamoto A., Kato I., Sakiyama F., Tsunasawa S.;
RT    "The primary structure of porcine aminoacylase 1 deduced from cDNA
RT    sequence.";
RL    J. Biochem. 112:737-742(1992).
CC    -!- FUNCTION: INVOLVED IN THE HYDROLYSIS OF N-ACYLATED OR
CC        N-ACETYLATED AMINO ACIDS (EXCEPT L-ASPARTATE).
CC    -!- CATALYTIC ACTIVITY: An N-acyl-L-amino acid + H(2)O = a fatty acid
CC        anion + an L-amino acid.
CC    -!- COFACTOR: ZINC.
CC    -!- SUBUNIT: HOMODIMER.
CC    -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC    -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; D13514; BAA02731.1; -.
DR    EMBL; X68564; CAA48565.1; -.
DR    PIR; JN0584; JN0584.
DR    PIR; S27010; S27010.
DR    MEROPS; M20.973.
DR    InterPro; IPR001261; ARGE_DAPE_CPG2.
DR    InterPro; IPR002933; Peptidase_M20.
DR    Pfam; PF01546; Peptidase_M20; 1.
DR    PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR    PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
```

```
KW    Hydrolase; Zinc; Acetylation.
FT    INIT_MET      0      0
FT    MOD_RES       1      1      ACETYLATION.
FT    CONFLICT    394    394      MISSING (IN REF. 2).
FT    CONFLICT    397    397      A -> T (IN REF. 2).
SQ    SEQUENCE  406 AA;  45216 MW;  1B0E3E69364OCBBE CRC64;
```

Alignment Scores:
```
Pred. No.:
Score:                    2.29e-07
Percent Similarity:       202.00       Length:              406
Best Local Similarity:    38.20%       Matches:             101
Query Match:              22.70%       Conservative:         69
DB:                       7.31%        Mismatches:          176
                          1            Indels:               99
                                       Gaps:                 18
```

US-10-014-896-1 (1-1509) x ACY1_PIG (1-406)

```
Qy      85 AGGAGGCGGGGAGGCCATCAAAGGGCGTCG----------------------CGAATCCCTTCTCAG 126
           ||||||||||||||||||||||                              |||||||
Db       5 ArgGlyGlyGluHisProSerValThrLeuPheArgGlnTyrLeuPheGlnIle----------- 21

Qy     127 TTCAGCAAAGAAGGAAACGGCGTCGGCAGTGAAAGGAGGCGGCTGAAAAGGTGGCCATCCAGAGTTCCA 186
           |||                ||                |||               |||||
Db      22 ------------------ArgThrValGlnProGluProAspTyrGlyAla------------ 32

Qy     187 ACAGTGACTTTTGACTCTGAGAAGTCCAATACTACCAGGCGTCCTGGCTGAGTTCGGAAAATAC 246
           |||                ||:||:::                ||
Db      33 AlaValAlaPheLeuGluGlnArgGlyAlaArg--------------GlnLeuGlyLeuGly 47

Qy     247 ATTCATAAAAGTCTTTCTCACACTGGTCAGCACCACCAGCTTTATCCAGACATGAAGTCGCGTGGAA 306
           ||||||                ||::                ||||||
Db      48 CysGlnLysVal-------------------------------GluValPro 55

Qy     307 GAGTATAGCCACCTGTTCACTATC---------------CAAGGCTCGGACACCCCAGGTTGCAGCCC 357
           ||                |||:::                ::::
Db      56 -----GlyHisValValThrValLeuThrTrpProGlyThrAsnProThrLeuSerSer 73

Qy     358 TACCTGCCTGCGATGGCCTCACTTTGATGTGGTGGTCGCCTGCCCCGTGAAGAAGGCTGGGAGGTGCCC 417
           |||                |||||||||||                |||
Db      74 IleLeuLeuAsnSerHisThrAspValValProValPheLysGluHisTrpSerHisAsp 93

Qy     418 CCATTCTCTTGGGTTG---GAGCGTGATGGCGTGCATCTATGGTGCGGGGGCACATGGACGAC 474
           |||||                |||                ||||||||||||||
Db      94 ProPheGluGlyPheLeuLysAspAlaAspSerGlyTyrIleTyrGlyArgGlyAlaGlnAspMet 113

Qy     475 AAGAACTCTGTGGATGGCAATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATC 534
           |||                ||:||::::                |||
Db     114 LysCysValSerIleGlnTyrLeuGluAlaValAlaArgArgLeuLysValGluGlyHisHis 133

Qy     535 CCCCGAAGATCTTTCTTCCATTTCTCCTGGGCCCCATGAGTGAGGAGGTCATCAGGG---ACAGG 591
           |||||                |||||||||||                :::
Db     134 PheProArgThrIleHisMetThrPheValProAspGluGluValAlgGlyGlyHisGlnGly 153

Qy     592 GCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGGGTCAGCCTTCAATTGTGGAC 651
           |||                |||                |||||||||| :::||
Db     154 MetGluLeuPheValLysArgProGlnIleAlaLeuArgAlaGlyPheAlaLeuAsp 173

Qy     652 GAGGGGGGCGTTCACCTTGGATGAATTCATTCCTAAACTTCAAGAAGCCCATCGCGCCTTGATT 711
           |||||                |||                |||
Db     174 GluGly-------------------------------------LeuAlaSerProThrAspAlaPhe 183

Qy     712 GCAGTC-----TCAGAGAAGGGTTCCATGAACCTCATGCGTCGCAAGTAAACATGACTTCA 765
           |||:::                |||||||:::
Db     184 ThrValPheTyrSerGluArgSerProTrpLeuArgValThrSerThrGlyLysPro 203

Qy     766 GGCCACTCTTCAGCTCCTCCAAAGGACGACAAGCAATTGGCATCCTTGCACTGCCTTGATT 825
           |||||                |||                |||
Db     204 GlyHisGlySerArgPheIleGluAspThr-------------AlaAlaGlu 216

Qy     826 CGATTGGAGCAGACACCAATGCCTATCAATATTTGGAAGGGACAGTGGTGACTGTATTG 885
           :::|||                |||||                ||
Db     217 LysLeuHisGlyLysVal--------------------------- 221
```

```
Qy   886 CAGCAACTGGCAAAATGAGTTTCCCTTCCCTGCAATAATAAATCCTGACCAACCCATGGCTA 945
                                 :::|||  ||||||  |||:||||:::
Db   222 ------------------------IleAsnSerIleLeuAla------------------ 227

Qy   946 TTTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGG 1005
             |||                                      :::
Db   228 PheArgGluLysGlyLysGlnArgLeuGlnSerAsnGlnLeuLysProGlyAlaVal--- 246

Qy  1006 ACCACCACGGCACTCACCATTATTCAAAGCAGGGTCAAGTTCATCCCCCAGTG 1065
             ||||||                                  |||
Db   247 ThrSerValAsnLeuThrMetLeuGlnLeuGlyLysValAlaTyrAsnValVal.ProLaThr 266

Qy  1066 GCCCAGGCCCACAGTCAACTTCCGGATTCACCCTGGCACAGTCCAAGAGGTCCTAGAA 1125
             |||         |||||||      :::|||       :::::
Db   267 MetSerAlaCysPheAspPheArgValAlaProAspSerAspProThrPro----- 286

Qy  1126 CTCACGGAGAAC------ATTGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGACTGGCC 1179
             ::::::                 |||         |||
Db   287 GlnLeuGlnSerTrpCysGlnAlaAlaAlaGlyIleGlyGluGlyGlyIleValAlaIleAlaTyrAsnValValProAlaThr 306

Qy  1180 TTTGACCCCCTCCCCGTCACGCCCTTGTGATCACAAGGCGTCCTGGCTACCAGCTGCTCCGC 1239
             |||:||||||||||||       :::|||||||       ||| |||
Db   307 TrpMetGluThrGlnValThrSerThrAspAspSerAspProThrTrpPro--- 322

Qy  1240 CAGACCGTACAGCTCCGTCTTCCCGGAGTCAATATTACTGCCCAGTTACTTCTATT--- 1296
             :::::: |||||||        :::
Db   323 AlaAlaAlaPheSerGlyValPheLysAspMetLysSerLeuAlaLeuGluLeuGluIleCysPro 342

Qy  1297 GGCAACACAGACAGCCGATTCTTTACAAAACCTCACCACTGGCATCATCCAGTTCCTGTTCTCTATT 1356
             |||       :::          |||      :::
Db   343 AlaSerThrAspAlaArgTyr-IleArgAlaAlaGlyValProAlaLeuGlyPheSerPr 362

Qy  1357 ATCTACACATACAGC 1369
             :::::
Db   362 oMetAsnHisThr 366
```

US-10-014-896-1 (1-1509) x Y457_METJA (1-410)

```
Alignment Scores:
Pred. No.:
Score:                 3.84e-07
Score:                 199.00          Length:        410
Percent Similarity:    38.08%          Matches:       99
Best Local Similarity: 20.71%          Conservative:  83
Query Match:           7.20%           Mismatches:    184
                       1               Indels:        112
                                       Gaps:          19
DB:
```

RESULT 3

Y457_METJA

```
ID   Y457_METJA     STANDARD;     PRT;     410 AA.
AC   Q57899;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein MJ0457.
GN   MJ0457.
OS   Methanococcus jannaschii.
OC   Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC   Methanocaldococcaceae; Methanocaldococcus.
OX   NCBI_TaxID=2190;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX   MEDLINE=96337999; PubMed=8668087;
RA   Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA   Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,
RA   Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA   Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA   Scott J.L., Geoghagen N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA   Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA   Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA   Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RL   "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL   jannaschii."
RL   Science 273:1058-1073(1996).
CC   -!- FUNCTION: COULD BE A PEPTIDASE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC   --------------------------------------------------
```

```
Db    57 TyrAsnIleIleAspLysTyrGlyIleGluArgProAsnIleValPheLysIleAspPhe 76

Qy   310 TATAGGCACCTG------------------ ... TTCACTATCCAA--- 333
              :::                                :::
Db    37 LysLysLysLeuMetGluTyrValGluSerTyrAsnIleGluAsnTyrThrLeuLysGlu 56

Qy   250 CATAAAGTCTTTCCTACAGTGGTCAGCACCACCAGCGTTATCCAGCATGAAGAGTCGTGGAAAG 309
             |||                                |||
Db    24 PheGlyGlyIleValGlyGlyLys---------------GluLysAlaGlyLeuTyrVal 36

Qy   202 -----------TCTGAGAAGTCCAATACTACAGCCCTGGCTGAGTTCGGAAAATACATT 249
                      ||||||                      :::
Db     4 IleGluGluAlaIleLysLeuGluSerAspLeuIleLeuArgIleAsnSerValAsnProSer 23

Qy   151 ATGAAAGAGGGGCCTGAAA--------------GGTGCCACCAGATTCAACACAGTGACTTTTAGC 201
             :::|||||||||||::::|                 ::|||    :::|||
Db    77 GlyArgAspLysThrLeuHis----------IleIleSerHisLeuAspPheVal.Pro--- 92

Qy   334 GGCTCGGACCCCCAGCTTGCAGCCCTTACCTGCTGATGGCGTCACTTTGA---GTGGTGCCTGCC 393
                   |||                        :::::||||||
Db    93 -----GluGlyAspIleSerLeuPheTrpGlyThrAsnProTyrGluProValIleLeuLysAsp 110

Qy   394 CCTGAAGAAGGC---------TGGGAGGTGCGCCCCAATTCTCTGTGGGTTGGAGCGTGAT 441
             ||||||                  :::       |||||||
Db   111 GlyLysIleTyrGlyIleArgGlySerGluAspAsnHisLysGlyIleValSerSerLeuLeu 130

Qy   442 GGCGTCACTCTATGGTCGGGGGGGCACACTGGACGACGAAGAACTCTGTGATGGCAATTACTGCAG 501
             |||      :::                   |||
Db   131 LeuLeuLysMetIlePheGluAsnAsnIleGluProLysTyrAsnLeuSerIleLeuePhe 150

Qy   502 GCCTTGGAGCTCCTGCTGATCAGGAAGAACATCCCCGAAGAGTCTTTCTTCATTTCTG 561
             |||:::::                       :::
Db   151 ValSerAspGlnLeuAspGly----SerGluTyrGlyTyrLeuLysTyrLeuLeuAsnAsn 168

Qy   562 GGCCATGATGAGGAGTCATCACCAGGGACAGGGGCGTCAGGACTCAGCCCTGCTACAGTCA 621
             ||||||                            :::
Db   169 --------------------PheGluAspLeuePheLysLysAspLeuIle 180

Qy   622 AGGGGCGTCCAGCTAGCCTTCATTGTGGACAGAGGGGGGCTTCATCATCTTGGATGAGTTCATT 681
             |||:::                           |||
Db   181 IleValProAspPheGlyThrProThrGluPheValGluIleGlyLysGlyLysIle 200

Qy   682 ------------CCTAACTTCAAGAAGCCCATCGCC---TTGATTGCAGTCTCACGAAAGGGTTCC 732
                         |||:::|||                  ||||||
Db   201 LeuTrpIleLysPheAsnIleLysGlyLysGlnSerGlyThrProGluGluGlyAsnGly 220

Qy   733 ATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCTGCCTCCTCCAAAGGAG 792
             :::                              ||||||
Db   221 LeuAsnLeuAlaAspLeuValAlaPheAsnPheAlaAsn------------ 232

Qy   793 ACAAGCAATTGGCAATCCTTGCACGTGCTGTCAGCCGATTGGAGCAGACACCAATGCCTATC 852

Qy   853 ATATTTGAAGCGGGACAGTCGTGCTGACGTGATTCGCAGCAACTGGCAAAATGAGTTTCCCTTC 912
```

Qy 233 ----------------------------GluLeuThrAsnGlyLeuThrTyrGluLysPheAspGlu 244
                                    :::   :::|||
Db 913 CCTGTCAATAATAATCCTGAGCAAACCATGGCTATTTGAACCACTTATAAGGAGGTTTATG 972

Qy 245 IleAsnSerIlePheLeuProGlyThrSerThrPheGluProThrIleLeuUysAsnLys 264
            :::||| ||| :::
Db 973 GAGAGAAATCCCTTAACCAAATGCAATAATCACGAACACCACGGCACTCACCATATTCAAA 1032

Qy 265 ValGluAsnPro--- 268
            ||||||||
Db 1033 GCAGGGGTCAAGTTCAATGTCAATCCCCCAGTGGCCCCAGTCAACTTCCGGATT 1092

Qy 269 -----------------AsnThrIleProGlyThrTyrValGluValValPheAspCysArgIle 283
                             |||||| ::: ||||| :::|||
Db 1093 CACCCTGGACAACAGAGTCCCAAGAGGTCCTAGAACTCACGAAGAACATTGTGGCT----- 1146

Qy 284 LeuProThrTyrLysIleGluGluValLeuGluPheIleAsnLysPheIleLysAsnPhe 303
                   |||||:::|||||||||
Db 1147 -----------------GATAACAGAGTGCCAG-------------------TTCCAT 1167

Qy 304 GluPheLysLysTyrIleLysHisTyrAspAsnSerIleLysAlaGluIleThrTyrGlu 323
                 |||||| :::||| :::
Db 1168 GTGTTGAGTGCCTTTGACCCCGTCCCGTCAGCCCTTCTGATGACAAGGCCTTGGGCTAC 1227

Qy 324 IleLeuArgLysSerGluAsnPro------------------------------------- 340
            :::||||| :::|||
Db 1228 CAGCTGCTCCGCCAGACGTACAGTCCGTCTCCCGGAAGTCAATAATTACTGCCCCAGTT 1287

Qy 341 LysGluLeuLysLysSerAlaIleLeuLysAsnVal---------------LeuAsnArgAspAlaLysLeu 357
            ::: ||| :::||| ||| :::
Db 1288 ACTTCTATTGGCAACACAGACAGCGGATTCTTTTACAAACCTCACCACTGACATCTACAGG 1347

Qy 358 CysGlyMetGlyGlyGlyThrArgValAlaAlaLaPheLeuArg----------TyrLys 372
                :::||| |||:::
Db 1348 TTCTACCCCATCTACATCACGCCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAAA 1407

Qy 373 GlyTyrAsnValAlaValTrpGlyIleGluGlyGluThrAlaHisGlnProAsnGluHis 392
            :::||| |||||
Db 1408 ATCTCAGTCCAAGGCCTATGAGACCCAAGTGAAATTCAICTTTGAGTTGAITCAG 1461

Qy 393 IleLysIleGluAspLeuValLysMetAlaGluValPheTyrGluIleLeuLys 410
            :::   :::|||||:::::::
Db 1461

RA   Iwaki K., Tanaka Y., Ohta T., Fukuda S., Kurimoto M.;
RL   Submitted (JUN-1993) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, Lung, and Skin;
RA   Strausberg R.;
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: INVOLVED IN THE HYDROLYSIS OF N-ACYLATED OR
CC   N-ACETYLATED AMINO ACIDS (EXCEPT L-ASPARTATE).
CC   -!- CATALYTIC ACTIVITY: An N-acyl-L-amino acid + H(2)O = a fatty acid
CC   anion + an L-amino acid.
CC   -!- COFACTOR: ZINC.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; L07548; AAA02852.1; -.
DR   EMBL; D14524; BAA03397.1; -.
DR   EMBL; D16307; BAA03814.1; -.
DR   EMBL; BC000545; AAH00545.1; -.
DR   EMBL; BC003023; AAH03023.1; -.
DR   EMBL; BC014112; AAH14112.1; -.
DR   PIR; S35709; S35709.
DR   MEROPS; M20.973; -.
DR   Genew; HGNC:177; ACY1.
DR   MIM; 104620; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Hydrolase; Zinc.
SQ   SEQUENCE 408 AA; 45885 MW; 293350CD7759826C CRC64;

US-10-014-896-1 (1-1509) x ACY1_HUMAN (1-408)

RESULT 4

ACY1_HUMAN
ID   ACY1_HUMAN   STANDARD;   PRT;   408 AA.
AC   Q03154;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Aminoacylase-1 (EC 3.5.1.14) (N-acyl-L-amino-acid amidohydrolase)
DE   (ACY-1).
GN   ACY1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93363640; PubMed=8357837;
RA   Mitta M., Kato I., Tsunasawa S.;
RT   "The nucleotide sequence of human aminoacylase-1.";
RL   Biochim. Biophys. Acta 1174:201-203(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=93352474; PubMed=8394326;
RA   Cook R.M., Burke B.J., Buchhagen D.L., Minna J.D., Miller Y.E.;
RT   "Human aminoacylase-1. Cloning, sequence, and expression analysis of
RT   a chromosome 3p21 gene inactivated in small cell lung cancer.";
RL   J. Biol. Chem. 268:17010-17017(1993).
RN   [3]
RP   SEQUENCE FROM N.A.

Alignment Scores:
Pred. NO.:   3.87e-06
Score: 185.50
Percent Similarity: 39.14%
Percent Identity: 39.14%
Best Local Similarity: 23.86%
Query Match: 6.72%
DB: 1

Length: 408
Matches: 89
Conservative: 57
Mismatches: 158
Indels: 69
Gaps: 13

Qy 319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTCGATGGTCACTTT 378
            ||| ::: |||||:::::::|||||| |||||||
Db   62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLeuAsnSerHisThr 81

Qy 379 GATGTGGTGCCTGCCCTGGGCCCTGAAGAAGGGCGACAAGAAGGCGACGACAGACT 435
            |||||||||||||||||| :::||| |||::: ||| :::
Db   82 AspAlaValProValPheLysGluHisTrpSerHisAspProProNeGluAlaPheLysAsp 101

Qy 436 CGTGAAGGGCGTCATCTATGGTCGGGCGCACACTGACGACAAGAAGAACTCTGATGGCATTA 495
            |||::: ||||||| |||||::: |||||||
Db 102 SerGluGlyTyrIleTyrAlaAlaArgLysValAlaGlnAspMetLysCysValSerIleGlnTyr 121

Qy 496 CTGCAGGCCTGGAGCCTCCTCGCTGATCAGGAGAAGTACATCCCCGAAGAGATCTTTCATT 555
            |||||| :::|||||| ||| :::|||||:::::::
Db 122 LeuGluAlaValArgLeuLysValGluGlyHisArgPheProArgThrIleHisMet 141

Qy 556 TCTCTGGGCCATGATGAGGAGGTCATCAGGGACACAGGGGCGCTCAGGAGGATCTCAGCCCCTGCTA 615
            :::|||||||||||||| :::||| |||
Db 142 ThrPheValProAspGluGluValGlyGly------------------------ 151

Qy 616 CAGTCCAAGGGCGTCCAGCTA------------------------GCCTTCATTGTGGACGAGGGGGC 660

CC   -!- CATALYTIC ACTIVITY: N-succinyl-LL-2,6-diaminoheptanedioate + H(2)O
CC   = succinate + LL-2,6-diaminoheptanedioate.
CC   -!- COFACTOR: COBALT OR ZINC (BY SIMILARITY).
CC   -!- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC   LYSINE FROM ASPARTATE SEMIALDEHDE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC   ---
DR   EMBL; AP001118; BAB12814.1; -.
DR   MEROPS; M20.UNA; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   TIGRFAMs; TIGR01246; dapE_proteo; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; FALSE_NEG.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Diaminopimelate biosynthesis; Lysine biosynthesis; Hydrolase; Cobalt;
KW   Zinc; Complete proteome.
SQ   SEQUENCE   375 AA; 41289 MW; 2BCAE6559P950644 CRC64;

US-10-014-896-1 (1-1509) x DAPE_BUCAI (1-375)

Alignment Scores:
Pred. No.:
Score: 180.00
Percent Similarity: 38.08%
Best Local Similarity: 22.80%
Query Match: 6.52%

Length: 375
Matches: 88
Conservative: 59
Mismatches: 151
Indels: 88
Gaps: 18

Qy   301 GTGGAAGAGTATAGGCACCTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 360
          |||||||| ||| ||||||||||||
Db   45 ValAsnAspThrLysAsnPheTrpAlaPheArgGlyThrGlyLysThrLeu---------- 61

Qy   361 CTGCTGATGGCTCACTTTGATGTGGTGCCT---GCCCCTGAGAGAGGCTGGAGGTGCCC 417
          || ||| ||||||| |||||||
Db   62 ThrPheAlaGlyHisThrAspValValProIleGlyGlnAspLysAspTrpGlnThrAsp 81

Qy   418 CCATTCTCTGGGTTGGAGGCGTGATGGGCGTCATCTATGGTCGGGGCGCACACTGACGACAAG 477
          |||||| ||| ||| ||||||||||||||:::
Db   82 ProPheGlnProValIleArgSerGlyTyrLeuPheGlySerAlaAspMetLys 101

Qy   478 AACTCTGATGCGATTACTGCAGGCCCTTGGGACTCCTCGTGCATCAGG----AAGTAC 531
          ||| |||||||||||||||||:::
Db   102 GlyAlaLeuAlaAlaMetIleThrAlaAlaGluArgPheValAsnLysPheProAsnHis 121

Qy   532 ATCCCCGGAAGATCTTTCTTCATTCTCTCGGGCCCATGATGAGGAGTCATCAGGACA-- 588
          |||||||||||||||
Db   122 LysGlyArgLeuSerPheLeuIleThr------SerAspGluUserSerAlaValAsp 139

Qy   589 GGGGCGTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGCGTCCAGCTAGCCTTC---ATT 645
          ||| ||| ||||
Db   140 GlyThrIleLysIleValGluTyrLeuMetSerLysArgAspMetIleAspTyrCysIle 159

Qy   646 GTGGACGAGGGGGCGTTC-----ATCTTGGATGAGATTTCATTCCTAACTTCAAGAAGCCC 699
          ||| ||| ||| ||| ||| ||||||| ||| |||
Db   160 ValGlyGluProSerSerThrAsnIleValGlyAspValIleLysAsn------------ 175

Qy   700 ATCGGCCTTGATTGCAGTCAGAGAGAAGGGTTCCATGAACCTGAGGTCGAAGTTAAACATG 759
          :: || |||||:::
Db   176 ------------------GlyArgArgGlySerIleThrAlaAsnIleThrIleTyrGly 189

Qy   760 ACTTCAGGCCACTCTTCACCTCTCCAAAGGAGACAACAAGCATTGGCATCCTTCCAGCTGCT 819

Alignment region (protein/DNA, left block):

Db   152 ---HisGlnGlyMetGluLeuPheValGlnArgProGlyGluPheHisAlaLeuArgAlaGly
          :::||||::::|||| |||
Qy   152 TTCATCTTGGATGGATTTCATTCCTAACTTCAAGAAGCCCATACGCCCTTGATTGCAGTC--- 170
Qy   661 PheAlaLeuAspGluGlyIleLeuAlaAsn----------ProThrAspAlaPheThrValPhe
          |||| ||| ||
Db   171 ---TCAGAGAAGGGTTCCATGAACCTGCTGCGAAGTAAACATGACTTCAGGCCACTCT 187
Db   188 TyrSerGluArgSerProTrpProValArgValThrGlyArgProGlyHisAla
          ||||||||::: ||| |||:::
Qy   718 ---TCAGAGAAGGGTTCCATGAACCTGCTGCGAAGTAAACATGACTTCAGGCCACTCT 207
Qy   775 TCAGCTCCTCCAAAGGAGACAAGCATTGGCAnCCTTGCAGCTGCTGCTCAGCCGAGTTGGAG
          ||||||||| ||| ::::||
Db   208 SerArgPheMetGluAspThr--------------AlaAlaGlyLysLeuHis 220
Db   835 CAGACACCAAATGCCTATCATATTTGGAAGCGGGACAGTGGTGAGTCGTATTGCAGCAACTG
          ::::||||||||||::::
Qy   835 ---------------ValValAsnSerIleLeuAlaPheArg 230
Qy   221 Lys--------------------------------------
          :::
Db   895 GCAAATGAGTTCCCTTCCCTGTCAATATAATATCCTGAGCAACCCATGGCTATTTGAACCA
          ||||||||| ::::
Db   231 GluLysGluTrp--------------GlnArgLeuGlnSerAsnProHisLeuLysGluGly 246
Qy   955 CTTATAACGAGGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAANTCAGGAGACCACACG
          ::::|||||:::
Qy   247 SerVal--------------------ThrSerVal 251
Db 1135 AAC----------ATTGTGGGCTGATAAACAGAGTGTCCAGTTCCATGTCTTGTTGAGTGTGCCC
          ::::
Db   292 SerTrpProCysGlnAlaLAaGlyLyGluGlyValThrArgLeuLeuProHeAlaGlnHisLysTrpHis 311
Qy 1015 GCACTCACCAATATTCAAAGCAGGGGTCAAAGTTCAATGTCATCCCCAGTGGCCCAGGCC
          |||||||||| :: ||| ||||||||||||||||||||
Qy   252 AsnLeuThrLysLeuGlnGluGlyValAlaValAlaTyrAsnValIleProAlaThrMetSerArgAla 271
Db 1075 ACAGTCAACTTCCGGATTCACCCTGGACAGAGAGTCCTAGACTCACGAAG
          ::: ||||||||| ||||||
Db 1075 ...
Qy   272 SerPheAspPheArgValAlaAlaProAspValAspPheIleGluGluGlnLeuGln 291
Qy 1249 CAGTCCGTCGTTCCCGGAAGTCAATATAATTACT---GCCCCAGTTACTTCTTATTGGCAACACA 1305
          ||| :::
Db   312 ProGlnValThrProThrAspSerAsnProThrPhrpLAlaAlaPheSerArg----- 329
Db   330 -----ValCysLysLysAspMetAsnLeuThrLeuGluProGlyIleMetProAlaAlaThr
          ::: |||||||
Qy   330 ---CTGCTGATGGCTCACTTTGAT 347
Qy 1306 GACAGCCGATTCTTTACAAAACCCTCACCACTGGGCGTCTCTACCCCAATCTACATA 1365
          |||:||||| |||||:::|| ||| |||:::
Db   348 AspAsnArgTyrIleuArgGlnValProAlaLeuGlyPheSerProMetAsnArg 367
Db 1366 CAGCCT----------GAAGACTTCAAAACGGATCCAT 1392
          ||| :: ||||||:::|||
Qy   368 ThrProValLeuLeuHisAspHisIleuAspProValArgLeuHis 380

```
Db    190 IleGlnGlyHisIleAlaTyrProAsp------------------------------------   198
              |||||        :::        |||
Qy    820 GTCAGCCGATTGGACGACGACACCA---------------ATGGCTATCATATTTGGAAGCGGG   867
              |||                     :::|||:|||
Db    199 -------------LeuIleAspAsnProIleHisLysGlyLeuProValIleLeu----------   212
Qy    868 ACAGTGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCCTGTCAATATAATC            927
              :::::|||                ||||:|||:        ::|||
Db    213 LysIleLeuSerIleLeuLysLysAspSerGlyAsnAspPhePheLeuProSerSerIleAsn    232
Qy    928 CTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATGGGAGAGAAATCCCTTA      987
              |||||        |||||||                        |||
Db    233 IleAlaAsn-----------------------------------------IleHisAlaGlyAsnGlyPhe   242
Qy    988 ACCAAATGCAATAAATCAGGACCCACCAGGGCACTCACCATAATTCAAAGCAGGGGTCAAGTTC   1047
              ::|                                |||
Db    236 -----------------------------IleHisAlaGlyAsnGlyPhe                  242
Qy   1048 --AATGTCATCCCCCCAGTGGGCCACAGTGCAACTTCACCCTGACACAG                  1104
              |||||||||||||||
Db    243 AsnAsnValIleProGlySerLeuPheValGlnPheAsnIleArgPheSerSerGluVal       262
Qy   1105 ACAGTCCAAGAGGTCCTGAGAACTCACGAGAAGAACATTGTGGCTGATAAACAGAGTTCAGTTC  1164
              ::|:|||        ||||||:|||:        |||
Db    263 SerGluLysHisIleSerGlnSerGlnIleValIleAsnIleLeuAsnSerAsnAspIleAsnTyr  282
Qy   1165 CATGTGTTGAGTGGTCTACCCCATCTACACATACAGCCTTGAGACTTCAAAGCACCATGGAGTC  1224
              |||        ||||||        |||
Db    283 SerIleGluTrpLeuPheSerGlyLysProPheIleThrLys-----LysGlyLeu--        299
Qy   1225 TACCAGCTGCTCCGCCAGACCGTACAGTCCGTCTTCCCGGAAGTCAATATTACTGCCCCA      1284
              |||:||||||::||||||        ::||        |||
Db    300 ------LeuIleAspThrValIleGlnSerIlePhe---TyrPheAsnLysLysLysPro       316
Qy   1285 GTT-----ACTTCTATTGGCAACAACAGCCGATTCTTTTACAAACCTCACCACTGGC        1338
              :||        |||||:        |||:|||||        :::
Db    317 IleLeuSerThrSerGlyGlyThrSerArgPheIleLeuAlaLeuMetCysSerGlu        336
Qy   1339 ATCTACAGGTTCTACCCCATCTACACATACAGCCTCGAAGACTTCAAAGCACCATGGAGTC     1398
              :::                        |||:||||||
Db    337 ValValGluGluGlyLeuValAsn----------------AsnThrIleHisLysVal        350
Qy   1399 AACGAGAAAATCTCAGTC     1416
              ||||||
Db    351 AsnGluCysValLysIle      356
```

RT   "Whole-genome random sequencing and assembly of Haemophilus Influenzae
RT   Rd.";
RL   Science 269:496-512(1995).
CC   -!- CATALYTIC ACTIVITY: N-succinyl-LL-2,6-diaminoheptanedioate + H(2)O
CC   = succinate + LL-2,6-diaminoheptanedioate.
CC   -!- COFACTOR: COBALT OR ZINC (BY SIMILARITY).
CC   -!- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC   LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC   -!-
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; U32695; AAC21776.1; -.
DR   MEROPS; M20.UNA; -.
DR   TIGR; HI0102; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   TIGRFAMs; TIGR01246; dapE_proteo; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Diaminopimelate biosynthesis; Lysine biosynthesis; Hydrolase; Cobalt;
KW   Zinc; Complete proteome.
SQ   SEQUENCE 377 AA; 41353 MW; C852C7550FFD4OC2 CRC64;
```
US-10-014-896-1 (1-1509) x DAPE_HAEIN (1-377)

Qy    352 CAGCCCTACCTTGCTGATGGCT---CACTTTGATGTGGTGCCTGCCCCTGAAGAAGGC---     405
              :::|||        :::        |||        ||| :::|||
Db     59 GluProValIleAlaIlePheAlaGlyHisThrAspValValProThrGlyAspGluAsnGln     78
Qy    406 TGGGAGGTGCCCCCATTCTCTGGGTTGGAGCCTGAGGGCGTGATCAGTCATCTATATGGTCGGCGGCACA     465
              |||||||||||        :::||||||||||||:||||||
Db     79 TrpSerSerProProPheSerAlaGluIleIleIleAspGlyMetLeuTyrGlyArgGlyAla     98
Qy    466 CTGGACGGACAAGAGAACTCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGCCTGATCAGG     525
              :::|||        ||: :::|||        |||||||
Db     99 AlaAspMetLysGlySerLeuAlaAlaMetIleValAlaAlaGlu                     113
Qy    526 AAGTACATC------CCCCGAAGANCTTTNCTTCANTTCTCTG-----------GGCCAT     567
              |||||||||        :::||| |||:|||||:|||:|||
Db    114 GluTyrValLysSerAlaAsnProAsnHisLysGlyThrIleAlaLeuLeuIleThrSerAsp     133
Qy    568 GATGAGGAGTCANTCANCAGGGACAGGGGCTCGAGAGAATCTCAGCCCTGCTACAGTCAAGGGGC     627
              :::|||        :::|||||| |||:||||||:::
Db    134 GluGluAlaThrAlaLysAspGlyThrIleHisValValGluThrLeuMetAlaArgAsp       153
Qy    628 GTCCAGCTAGCCTTC---ATTGTGGACGAGAGGGGGGCCTTCANCTTGGATGATTTCATTCCT     684
              :::|||||| |||        :::|||        |||
Db    154 GluLysIleThrTyrCysMetValGlyGlu-----------------------------Pro   164
Qy    685 AACTTCAAGAAGCCCCATCGCC---TTGATTGCAGTCTCAGAGAAGAGGGTTCCATGAACCTC     741
              :::||| :::|||        :::|||||||:::
Db    165 SerSerAlaAlaGluGlyArgLeuGlyAspValValLysAsnGlyArgArgGlySerIleThrGly     184
Qy    742 ATGCTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCTGCTCCTCCAAAGAGACAAGCANTT     801
              |||||| :::|||
Db    185 AsnLeuTyrIleGlnGlyIleGlnHisValAlaTyrPro------------------------   198
```

```
RESULT 6
US-10-014-896-1

SEQUENCE FROM N.A.

Alignment Scores:
Pred. No.:
Score:                5.37e-05
Percent Similarity:      170.00          Length:           377
Best Local Similarity:    37.53%         Matches:           86
Query Match:              22.57%         Conservative:      57
                           6.15%         Mismatches:       132
                              1           Indels:          106
                                          Gaps:             19
DB:    US-10-014-896-1 (1-1509) x DAPE_HAEIN (1-377)
```

```
Qy    802  GGCATCCTTGCGACTGCTGTCGTCCGGATTGGAGCAGGAGACACCAAATGCCTATCATATTTGGA  861
                ||| ::|||          |||          ||:::
Db    199  ----------------------HisLeuAlaGluAsnProIle---------------------  205

Qy    862  ACGCGGGACAGTGTGACTGTTGACTGTTGCAGCCAACTG--------------GCAAAT       900
                |||||::::||||||||            |||:::|            |||
Db    206  ---HisLeuAlaLeuPheLeuGlnGluLeuThrThrTyrGlnTrpAspLysGlyAsn        224

Qy    901  GAGTTCCCTTCCCTGTCAATATAATCCTGAGCCACCATGGCTATTGTGAACCACTTATA      960
                |||||||              ||:::||              ||:::|||
Db    225  GluPhePheProProThrSerLeuGlnIleLeuAsn------------------------  236

Qy    961  AGCAGGTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGAACCACCACGGCACTC     1020
                |||||||||              |||              :::
Db    236  ------------------------------------------------------------  236

Qy   1021  ACCATAATTCAAAAGCAGGGGTC---AAGTTCAAATGTCATCCCCCAGTGGCCCAGGCCACA     1077
                |||||||              |||||||||              ||:::
Db    237  ----------IleHisAlaGlyThrGlySerAsnAsnValIlePro-----AlaGluLeuTyr  252

Qy   1078  GTCAACTTCCGGATTCACCCTGGACAGACAGTC-----------------CAAGAGGTC     1119
                |||||||              |||||||||||              |||:::
Db    253  IleGlnPheAsnLeuArgTyrCysThrGluValThrAspGlnIleIleLysGlnLysVal     272

Qy   1120  CTAGAACTCACGAGAGAACATTGTGGCTGATAAACAGTTCCAGTTCCATGTGTTGAGTGCC     1179
                ||:::|||||              ||:::||||||              ||:::
Db    273  AlaGluMetLeuGluLysHisAsnLeuLysTyrArgIleGluTrpPheAsnLeu-----Ser  290

Qy   1180  TTTGACCCCTCCCGTGCCTTCGATGACAAGGCCTTGGGTACCACCTGCTCCCGC     1239
                :::|||||              |||||||              ||:::|
Db    291  GlyLysProPheLeuThrLysProGly--------------------LysLeuLeuAsp     303

Qy   1240  CAGACCGTACAGTGCCGTCCTGCTCTCCCGGAGTCAAATATTACTGCCCCCAGTT--ACTTCTATT     1296
                ||| :::|||              |||:::||||||              ||:::
Db    304  SerIleThrSerAlaIleGlyLeuThrIleGlyIleThrProLysAlaGluThrGlyLysGly     323

Qy   1297  GGCAACACAGACAGCCGATTCTTTACAAACCTCACCACTGGCCATCTCAGGTTCTCACCCC     1356
                ||| :::|||              |||:::|||||||              ||:::
Db    324  GlyThrSerAspGlyArgPheLeuAlaLeuMetGlyAlaGlnValValGluProPheGlyPro     343

Qy   1357  ATCTACATACAGCCTGAGACACTTCAAACGCCATCCATGGAGTCAACGAGAAAATCTCAGTC     1416
                ||||||              |||||||              ||:::|
Db    344  -----SerHisIleHisLysValAsnGluCysValSerVal     357

Qy   1417  CAA     1419
                :::
Db    358  Glu     358
```

RT   "The essential Escherichia coli msgB gene, a multicopy suppressor of
RT   a temperature-sensitive allele of the heat shock gene grpE, is
RT   identical to dapE.";
RL   J. Bacteriol. 174:5258-5264(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=97349980; PubMed=9205837;
RA   Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA   Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA   Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA   Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasundaram S.,
RA   Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA   Yamagata S., Horiuchi T.;
RT   "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT   - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT   analysis of its sequence features.";
RL   DNA Res. 4:91-113(1997).
CC   -!- CATALYTIC ACTIVITY: N-succinyl-LL-2,6-diaminoheptanedioate + H(2)O
CC   = succinate + LL-2,6-diaminoheptanedioate.
CC   -!- COFACTOR: COBALT OR ZINC.
CC   -!- PATHWAY: FIFTH STEP IN THE BIOSYNTHESIS OF DIAMINOPIMELATE AND
CC   LYSINE FROM ASPARTATE SEMIALDEHYDE.
CC   -!- MISCELLANEOUS: DAPE/MSGB IS A MULTICOPY SUPPRESSOR OF A
CC   TEMPERATURE-SENSITIVE ALLELE OF THE HEAT SHOCK GENE GRPE.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; S41760; AAB22798.1; -.
DR   EMBL; X57403; CAA40665.1; -.
DR   EMBL; D90875; BAA16346.1; -.
DR   EMBL; AE000334; AAC75525.1; -.
DR   EMBL; D90876; BAA16350.1; -.
DR   PIR; A42958; A42958.
DR   PIR; A42959; A42959.
DR   MEROPS; M20.UNA; -.
DR   EcoGene; EG10208; dapE.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   TIGRFAMs; TIGR01246; dapE_proteo; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Diaminopimelate biosynthesis; Lysine biosynthesis; Hydrolase; Cobalt;
KW   Zinc; Complete proteome.
SQ   SEQUENCE   375 AA; 41269 MW; 181F43D05E88C20E CRC64;

RESULT 7
DAPE_ECOLI
```
ID   DAPE_ECOLI     STANDARD;      PRT;   375 AA.
AC   P24176;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-MAR-1992 (Rel. 21, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Succinyl-diaminopimelate desuccinylase (EC 3.5.1.18) (SDAP).
GN   DAPE OR MSGB OR B2472.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92355499; PubMed=1644752;
RA   Bouvier J., Richaud C., Higgins W., Bogler O., Stragier S.;
RT   "Cloning, characterization, and expression of the dapE gene of
RT   Escherichia coli.";
RL   J. Bacteriol. 174:5265-5271(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92354498; PubMed=1644751;
RA   Wu B., Georgopoulos C., Ang D.;
```

```
Alignment Scores:
Pred. No.:
Score:              0.000163                 Length:          375
Percent Similarity:  163.50                  Matches:          85
Best Local Similarity: 36.60%                Conservative:     50
Query Match:         23.10%                  Mismatches:      134
DB:                  5.928                   Indels:           99
                         1                   Gaps:             18
```

```
Qy   373 CACTTTGAATGTGGTGGTGCCTGCCCCCTGAA---GAAGGCTGGGAGGTTGCCCCCATTCTCTGG 429
         |||  ||||||||||||||||  :::  ||||  :::  ||||||||||||||||
Db    66 HisThrAspValProGlyProGlyAspAlaAspArgTrpIleAsnProProPheGluPro 85

Qy   430 TTGGAGCGTGATGGGTGCGTCAATCTATGGTCGGGGCCACACTGGACGACGAGAACTCTGGTATG 489
         |||||||||||||||||||||:::::||||||||||||||||||||
Db    86 ThrIleArgAspGlyMetLeuPheGlyArgGlyAlaAlaAlaAspMetLysGlySerLeuAla 105

Qy   490 GCATTACTGCAGGCCTTGGAGCCTCCTGCTGATCAGG-----AAGTACATCCCCGGAGA 543
         |||:::|||  |||||||||  |||  :::  |||
Db   106 AlaMetValValAlaLeuGluArgPheValAlaGlnHisProAsnHisThrGlyArgLeu 125

Qy   544 TCTTTCTTCATTTCTTCTGGGCCATGATGAGGAGTCATCAGG---ACAGGGGCTCAGAGG 600
         ||:::|||  ||||||||||||||||:::
Db   126 AlaPheLeuIleThr-----SerAspGluAlaSerAlaHisAsnGlyThrValLys 143

Qy   601 ATCTCAGACCCTGCTACAGTCAAAGGGGGTCCAGCTAGCCTTC--ATTGTGGACGAGGGG 657
         |||:::||| |||||| :::||  :::||| :::
Db   144 ValValGluAlaLeuMetAlaArgAsnGluLeuAspTyrCysLeuValValGlyGluPro 163

Qy   658 GGCTTCATC-----TTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCTTTGATT 711
         |||:::      |||  |||  :::  |||
Db   164 SerSerIleGluValValGlyAspValValLysAsn---------------------- 175

Qy   712 GCAGGTCTCAGAGAGAGGGTTCCATGAACCCTCATGCTGCAAGTAAACATGACTTCAGGCCAC 771
         :::|||||||||  |||||||||:::|||
Db   176 ------GlyArgArgGlySerLeuThrCysAsnLeuThrIleHisIleGlyValGlnGlyHis 193

Qy   772 TCTTCAGTCTCTCCAAAGGAGACAAGCATTGGCATCCTTGCCAGCTGCGTCTCAGCCGATTG 831
         |||  |||  :::  |||||||  :::|||
Db   194 ValAlaTyrPro------------------------------------------HisLeu 199

Qy   832 GAGCAGACACCAATG--------------CCTATCATATTTGGAAGCGGGACAGTGGTGACT 879
         ||||||  :::||| |||:::
Db   200 AlaAspAsnProValHisArgAlaAlaProPheLeu------AsnGluLeuValAla 216

Qy   880 GTATGCAGCAACTGGCAAAATGAGTTTCCCTTCCTGTCAATATAATCCTTGAGCAACCCA 939
         :::|||||||||||||||
Db   217 IleGluTrpAspProGlyLysAsnGluPhe---PheProAla------------ 228

Qy   940 TGGCTATTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTACCAATGCAATA 999
         :::|||||||
Db   228 ---------------------------------------------- 228

Qy  1000 ATCAGGACCACCACGGCACTCACCATTATTCAAAGCAGGGGTC--AAGTTCAATGTCATC 1056
         |||  :::  ||||||
Db   229 -----ThrSerMetGlnIleAlaAsnIleGlnAlaGlyThrGlySerAsnValIle 246

Qy  1057 CCCCCAGTGGCCCAGGCCCACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAG 1116
         |||||||||||||||||||
Db   247 ProGlyGluLeuPheValGlnPheAsnPheArgPheSer----------------- 259

Qy  1117 GTCCTAGAACTCACGAAGAACATTGTGGCTGATAAACAGAGTCCAGTTCCAGTGTTGAGT 1176
         |||||||  :::  |||||||
Db   260 ---ThrGluLeuThr-----------AspGluMetIleLysAlaGlnInValLeuAla 273

Qy  1177 GCCTTGACCCCTCCCCGTCAGCCCTTCTGATGACAAGGCCTTGGGCTACCAG----- 1230
         |||  :::  |||
Db   274 LeuLeuGluLysHisGlnLeuArgTyrThrValAspTrpProLeuSerGlyGlnProPhe 293

Qy  1231 ---------------CTGCTCCGGCCAGACGTACAGTCCGTCTCCGGAGTCAATATT 1275
         |||  :::  |||  :::
Db   294 LeuThrAlaArgGlyLysLeuValValAsnAlaValGluHisTyrAsnGlu 313

Qy  1276 ACTGCCCCA---------GTTACTTCTATTTGGAACACAGACAGCGGATTCTTTTACAAACCTC 1329
         |||:::||||:::||  |||||||
Db   314 IleLysProGlnLeuLeuThrThrGlyIleGlyThrSerAspGlyArgPheIleAlaArgMet 333

Qy  1330 ACACAGTGGCATCTACAGGTTCTACCCATCACATACAGCTCATGAGACTTCAAACGCATC 1389
         |||::::::
Db   334 GlyAlaGluIleValValGluLeuGlyProValAsn-----------AlaThrIle 347
```

```
Qy  1390 CATGGAGTCAACGAGAAAATCTCA 1413
         |||:::||||||||  :::::
Db   348 HisThrAspValIleAsnGluCysValAsn 355
```

RESULT 8
YB9X_YEAST

```
ID   YB9X_YEAST       STANDARD;      PRT;    878 AA.
AC   P38149;
DT   01-OCT-1994 (Rel. 30, Created)
DT   01-OCT-1994 (Rel. 30, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical 98.1 kDa Trp-Asp repeats containing protein in
DE   PAF1-MRPL27 intergenic region.
GN   YBR281C OR YBR2018.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94378722; PubMed=8091861;
RA   Holmstroem K., Brandt T., Kallesoe T.;
RT   "The sequence of a 32,420 bp segment located on the right arm of
RT   chromosome II from Saccharomyces cerevisiae.";
RL   Yeast 10:47-62(1994).
CC   -!- SIMILARITY: CONTAINS 6 WD REPEATS (TRP-ASP DOMAINS).
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X76053; CAA53644.1; -.
DR   EMBL; Z36150; CAA85245.1; -.
DR   PIR; S44543; S44543.
DR   PIR; S39137; S39137.
DR   MEROPS; M20.UPA; -.
DR   SGD; S0000485; YBR281C.
DR   InterPro; IPR002933; Peptidase_M20.
DR   InterPro; IPR001680; WD40.
DR   Pfam; PF00400; WD40; 4.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PRINTS; PR00320; GPROTEINBRPT.
DR   SMART; SM00320; WD40; 3.
DR   PROSITE; PS00678; WD_REPEATS_1; 2.
DR   PROSITE; PS50082; WD_REPEATS_2; 2.
DR   PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW   Hypothetical protein; Repeat; WD repeat.
FT   REPEAT       18     57       WD 1.
FT   REPEAT       68    107       WD 2.
FT   REPEAT      235    274       WD 3.
FT   REPEAT      282    322       WD 4.
FT   REPEAT      362    405       WD 5.
FT   REPEAT      608    651       WD 6.
SQ   SEQUENCE   878 AA;  98070 MW;  A24C10B94E94C344 CRC64;
```

Alignment Scores:

```
Pred. No.:                           0.000378
Score:                             160.00
Percent Similarity:                 36.98%
Best Local Similarity:              20.79%
Query Match:                         5.79%
```

```
                    Length:           878
                    Matches:           95
                    Conservative:      74
                    Mismatches:       200
                    Indels:            88
                    Gaps:              15
```

US-10-014-896-1 (1-1509) x YB9X_YEAST (1-878)

```
Qy    34 GTGGCTAAGCTGCTCCTAGTTTCCCTACCGTCTCCGATGATGGGCCCCGAGGAGGACGGGG 93
         ::|||||  ||||||  |||  |||:::
Db   397 IleThrSerLeuSerAlaValProLeuSerSerAsnSerProIleAsnAlaSerSer 416
```

```
RESULT 9
PEPV_LACDL

ID   PEPV_LACDL     STANDARD;     PRT;   470 AA.
AC   P45694;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Xaa-His dipeptidase (EC 3.4.13.3) (X-His dipeptidase) (Aminoacyl-
DE   histidine dipeptidase) (Carnosinase).
GN   PEPV.
OS   Lactobacillus delbrueckii (subsp. lactis).
OC   Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC   Lactobacillus.
OX   NCBI_TaxID=29397;
RN   [1]
RP   SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC   STRAIN=DSM 7290 / WS87;
RX   MEDLINE=95093606; PubMed=7528082;
RA   Vongerichten K., Klein J., Matern H., Plapp R.;
RT   "Cloning and nucleotide sequence analysis of pepv, a carnosinase gene
RT   from Lactobacillus delbrueckii subsp. lactis DSM 7290, and partial
RT   characterization of the enzyme.";
RL   Microbiology 140:2591-2600(1994).
-!- FUNCTION: HAS ACTIVITY AGAINST BETA-ALANYL-DIPEPTIDES INCLUDING
-!- CARNOSINE (BETA-ALANYL-HISTIDINE).
-!- CATALYTIC ACTIVITY: Hydrolysis of Xaa-|-His dipeptides.
-!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z31377; CAA83252.1; -.
DR   MEROPS; M20.004;
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Hydrolase; Dipeptidase; Metalloprotease.
SQ   SEQUENCE   470 AA;  51990 MW;  488117B4F33E4AB0 CRC64;
```

Alignment Scores:

Alignment No.:	9	
Pred. No.:		
Score:	149.00	
Percent Similarity:	32.41%	
Best Local Similarity:	20.46%	
Query Match:	5.39%	

Length:	470	470
Matches:	89	89
Conservative:	52	52
Mismatches:	146	146
Indels:	148	148
Gaps:	16	16

DB: 0.00208

```
US-10-014-896-1 (1-1509) x PEPV_LACDL (1-470)
```

(Query/Database nucleotide–protein alignment, Qy positions 94–1053, Db positions 141–813)

```
Qy  133 AAAGAGGAACGCGTCGGATGAAAAGAGGCG---CTGAAAGGTGCCATCCAGATTCCAACA 189
            |||||||                    ||||||          :::::
Db    6 LysGluLeuLeuAlaGlnLysLysLysAsp...AlaIleLeuLysLeuGluGluLeuIleAla 25

Qy  190 GTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGCCTGAAAATATACATT 249
            :::                 |||
Db   26 IleAspSerSerGluAspLeuGluAsnAlaThrGluAlaThrTyrProValGlyLys------ 43

Qy  250 CATAAAGTCTTTCCTTACAGTGGTCAGCACCAGCACCAGCTTTATCCAGCAT---------- 294
                          |||
Db   44 ------------GlyProValAspAlaMetThrLysPheLeuSerPheAlaAlaLysArgAspGly 60

Qy  295 -----GAAGTCGTGGAAGAGTATATAGCCACCACTGTTCACTATCCAAGGCTCGGACCCC 345
             |||
Db   61 PheAspThrGluAsnPheAlaAsnTyrAlaGlyArgValAsnPheGlyAlaGlyAsnLys 80

Qy  346 AGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGTGGTGCCCCTGAAGAAGGC 405
             |||                    ::::
Db   81 ArgLeuGly--------IleIleGlyHisMetAspValValProAla---GlyGlyGly 96

Qy  406 TGGGAGGGTGCCCCCCATTCTCT---GGGTTGGAGCCGTGATAGGCGTCATCTATGGTCGGGGGC 462
             |||                     ::::
Db   97 TrpThrArgAspProPheLysMetGluIleGluAspGluAlaThrGlyIleGluTyrGlyArgGly 116

Qy  463 ACACTGGACGACAAGAACACTCTGTGATGGCATTACTGCAGGCCGTTGGAGTCCTGCTGATC 522
             |||||||||
Db  117 SerAlaAspAlaSerGlyProSerLeuThrAlaTyrTyrGlyMetLeuLeuLeuLysGlyGlu 136

Qy  523 AGGAGAGTACATCCCCGAGAGATCTTTCTTCATTTCTCTGGGCCCAATGATGAGGAGTCATCA 582
             :::               |||||||:::||
Db  137 AlaGlyPheLysProLysLysLysIleAspPheValLeuGlyThrAsnGluGluThrAsn 156

Qy  583 GGGACAGGGGCTCAGAGGATCTCAGCCCCTGCTACAGTCAAGGGCGGTCCAGCTAGCCTTC 642
             :::               ||||||
Db  157 TrpValGly------------------------SerProAspAlaGlu 159

Qy  643 ATTGTGGACGAGGGGGGCTTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATC 702
             :::               :::
Db  160 ---IleAspTyrTyrTyrLeuLysHisGluProThrProAsp 171

Qy  703 GCCCTTGATTGCAGTCTTCAGAGAGAAGGGTTTCCATGGCTGCCAAGTAAAACATGACT 762
             :::
Db  172 IleValPhe------GluPheSerPhe 174

Qy  763 TCAGGCCCACTCTTCAGCTCTCCAAAGGAGCACAAGCCATTGGCATCCTTGTCAGCTGCTGTC 822
             :::               ---SerProAspAlaGlu---
Db  175 ---SerProAspAlaGlu--- 179

Qy  823 AGCCGGATTGGGAGCAGACACCAAATGCCTATACATATTTGGAAGCGGGGACCAGTGGTGACTGTA 882
             |||||||||          :::   |||
Db  180 ---TyrProIleIleAsnGlyGlyValGlnIlePheAsnThrLeu 192

Qy  883 TTGCAGCAACTGGCAAATGAGTTTCCCTTGTCCTGTCCAATATAAATCCTGAGCAACCCATGG 942
             :::
Db  193 ---GluPheSerPhe--- 196

Qy  943 CTATTTGAACCACTTATAAGCCAGGTTTATGGAGAGAATCCCTTAACCAATGCAATAAATC 1002
             :::|||
Db  197 ---LysAsnAspSerThr 201

Qy 1003 AGGACCACCACGGCACTCACCATAATTCAAGCAGGGGGTCAAGTTCAAGTGTCATCCCCCA 1062
             :::        |||||||||:::
Db  202 LysGlyAspTyrValLeuAspLysPheAlaGlyIleLeuAlaThrAsnValThrProGln 221

Qy 1063 GTGGGCCCAGGCCCAGGATTCCAACCTTCCGGACAGACAGTCCAAGAGGTCCTA 1122
             |||        ||||||||||:::
Db  222 ValThrArgAlaThrIleSer---GlyProAspLeuGluLeuAlaValLys------ 236

Qy 1123 GAACTCACGAAGACAAACATTGTGGCTGATAAACAGAGTTCAGTTCCATGTGTTCAGTGCCTTT 1182
             :::::|||:::||||||
Db  237 LeuAlaTyrGluSerPheLeuAlaAspLysGluLeuAsp------GlySerPhe 252

Qy 1183 GACCCCCTCCCCGTTCCGTATAGCCCTTCTGATGACAAGAGGCCCTTGGGGCTACCAGCTGGTCCGGCCAG 1242
```

```
Db  253 Glu-----------------------IleAsnAspGluSerAlaAspIleValLeuGlyGln 266
             :::                   ::::||||:::::
Qy 1243 ACCGTACAGTCCGTGCTTCCCGGAAGTCAAATATTACTGCCCAGTTACTTCTATTGGCAAC 1302
             :::        |||:::||||
Db  267 GlyAlaHisAlaSerAlaProGlnVal------------------------Gly 276
Qy 1303 ACAGACAGCGCGATGTCTTTACAAACGCTCCACCACTGGCATCACTCACAGGTTC------- 1350
             :::||||          |||          |||
Db  277 LysAsnSerAlaThrPheLeuAlaLeuPheLeuAspGlnTyrAlaLaPheAlaGlyArgAsp 296
Qy 1351 -----------------TACCCCARTCTACACATACAGCGCCTGAAGACTTC 1380
                         :::  ||||||||||
Db  297 LysAsnPheLeuHisPheLeuAlaGluValValGluHisGluAspPhePhe 311
```

RESULT 10
YGEY_ECOLI
ID   YGEY_ECOLI   STANDARD;   PRT;   403 AA.
AC   Q46805;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ygeY.
GN   YGEY OR B2872 OR Z4211 OR ECS3745.
OS   Escherichia coli.
OS   Escherichia coli O157:H7.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562, 83334;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=97426617; PubMed=9278503;
RA   Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA   Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA   Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA   Mau B., Shao Y.;
RT   "The complete genome sequence of Escherichia coli K-12.";
RL   Science 277:1453-1474(1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX   MEDLINE=21074935; PubMed=11206551;
RA   Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA   Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA   Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA   Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA   Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA   Welch R.A., Blattner F.R.;
RT   "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL   Nature 409:529-533(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=O157:H7;
RX   MEDLINE=21156231; PubMed=11258796;
RA   Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA   Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA   Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA   Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT   "Complete genome sequence of enterohemorrhagic Escherichia coli
RT   O157:H7 and genomic comparison with a laboratory strain K-12.";
RL   DNA Res. 8:11-22(2001).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A.
DR   EMBL; U28375; AAA83053.1; -.

```
DR   EMBL; AE000370; AAC75910.1; -.
DR   EMBL; AE005517; AAG58001.1; -.
DR   EMBL; AP002563; BAB37168.1; -.
DR   MEROPS; M20.UPA;.
DR   EcoGene; EG13055; yeeY.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
KW   Hypothetical protein; Hydrolase; Metalloprotease; Complete proteome.
SQ   SEQUENCE   403 AA;  44804 MW;  8793701224403BB6 CRC64;
```

```
Alignment Scores:
Pred. No.:                              Length:        403
Score:                  0.00305         Matches:       83
                        146.50          Conservative:  74
Percent Similarity:     34.89%          Mismatches:    166
Percent Identity:       18.44%          Indels:        127
Best Local Similarity:  5.30%           Gaps:          16
Query Match:            1
```

```
US-10-014-896-1 (1-1509) x YGEY_ECOLI (1-403)

Qy    106 GCGTCGCGAATCCCTTCTCAGTTC------------...AGCAAAGAGGAACGCGTCGCGATG 153
          |||    ::|                                      :::      |||
Db      2 AlaLysAsnIleProPheLysLeuIleLeuGlyLysAlaLysAspThrGlnAlaAspMet 21

Qy    154 AAAGAGGCGTGAAAGGTGCCATCCAACAGTGACTTTTAGCTCTGAAGTCC 213
          |||:::           ||||||            :::
Db     22 ThrArgPheLeuArgAspMetValAlaIleProSerGluSer---CysAspGluLys--- 39

Qy    214 AATACTACAGCCCTGGCTCGAGTCGGAAAATACAATTCAATAAAGTCTTCCTACAGTGGTC 273
          :::      ::                      :::||||
Db     40 --------ArgValValHisArgIleLysGluLeuMetGlu 50

Qy    274 AGCACCAGCTTATCCCAGCATGAAGTGCGTGAAGAGTATAGCCACCTGTTCACTATCCAA 333
          ::||        :::||||       :::  ||||||
Db     51 LysValGlyPheAspLysValGlu---IleAspProMetGlyAsnValLeuGlyTyrIle 69

Qy    334 GGCTCGGAACCCAGCGTTGCAGCCCTACCTGCTGATGGCTCCACTTTGATGTGGGCCATGAT 393
          |||     |||            :::||||||        |||
Db     70 GlyHisGlyProArgLeu--------ValAlaMetAspAlaHisIleIleAspThrValGlyIle 87

Qy    394 CCTGAA------GAAGGCTGGGAGGTGCCCCATTCGTCTGGGTTGGAGGCGTGGATGGCGTCATC 450
          ::||         |||:::      ||||||:::              ||||
Db     88 GlyAsnIleLysGlyAsnThrPheAspProTyrGlyGlyMetGluThrAspGluLeuIle 107

Qy    451 TATGGTTCGGGGCCACACTGCGACGACGACAAGAACTCTGTGAGCAATTACTGCAGGCCTTGGAG 510
          ||||||||:::||||            :::||||
Db    108 GlyGlyArgGlyThrSerAspGlnLeuGlyMetAlaSerMetValTyrAlaLeuGlyLys 127

Qy    511 CTCCTGCTGATCAGGAAGTACATCCCCGAAGATCTTTCTTTCATTTCTCTGGGCCATGAT 570
          ||||:::::||                      |||::
Db    128 IleIleLys-----------------------AspLeuGlyLeuGlu 135

Qy    571 GAGGAG-----------------TCATCAGGACAGGGGCTCAGAGGATCTCAGCCCGCTACAG 618
          :::|||                   :::||||||
Db    136 AspGluTyrThrLeuLeuValThrGlyThrValGlnGlu------------------Asp 150

Qy    619 TCAAGGGGGGTCGTCCAGCTAGCCTTCATTGTTGGACGAGGGGGGCTTCACTGGATGATTTC 678
          ||||||||:::||||            :::||||
Db    151 CysAspGlyLeuCysTrpGlnTyrIleIleLeuGlnSerGly---IleArgProGluPhe 169

Qy    679 ATTCCTAACTTCAAGAGAAGCCCATGCCTCAGTCTCAGAGAAGGGTTCCATGAAC 738
          ::                :::||||            |||
Db    170 ValValSerThrGlnProThrAspCysGlnValThrArgGlyGlnAlaArgMetGlu 189

Qy    739 CTCATGCTGCAAGTAAAACATGACTTCAGGCACCACTCTTCAGCTCCTCCAAAAGGAGACAAGC 798
          ::::||||     :::||||||            |||
Db    190 IleArgIleAspValGlnLeuValSerCysHisGlySerAlaProAlaGluAlaArgGlyValAsnGluAsn 209

Qy    799 ATTGGACATCCTTGCCAGTCTGCTGTCAGCCGGATTGGAGCGAGCAGACACCAACGCCTATCCATATTT 858
          :::                                    |||||||
Db    210 -------------------------------------------------AlaIlePhe 212

Qy    859 GGAAGCGGGGACCAGGTGGTGGACGTGTATTGCAGCAAACTGGCAAATGAGTTTCCCTTCCCTTGTC 918
          |||:::::                                           :::
Db    213 LysMetGlyProIleLeuGlyIleLeuGlyInLeuSerGlnAsnArgGlnGlyTyr------ 230

Qy    919 AATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATTAAGCAGGTTTATGGAGAGA 978
          :::||||||                  ||||:::
Db    231 --------------------------------------------------IleArgProGluLys 236

Qy    979 AATCCCTTACCAATGCAATAATCAGGACACCACCACGGCGACTCACCATTATCAAAGCAGGG 1038
          |||      :::                       |||::
Db    237 Gly---------------------------------ThrLeuThrValSerGluIlePhePheThrSer 248

Qy   1039 GTCAAGTTCAATGTCATCCCCCAGTGGCCCCAGCAGTGCAAGTCAACTTCCGGATTCACCCT 1098
          |||:::                    :::            |||::
Db    249 ProSerArgCysProSerAlaLeuAlaAspSerCysSerAlaValSerIleAspArgGlnGlyLeuThrArgGln 268

Qy   1099 GGACAGACAGTCCAAGAGGGTCCTAGAACTCACGAAGAACACATTGTGTGGCTGATAACAGAGTC 1158
          |||:::::                    |||||||
Db    269 GlyGluThrTrpGlu---------------------------------------------------- 273

Qy   1159 CAGTTCCATGTGTTGAGTGCCTTTGACCCCCTCCCCGTGCAGCCCCTTCTGATGACAAGGCC 1218
          :::||                       :::          :::||||
Db    274 -------------------------------------------GlyAlaLeuAspGluIleLeuArgAlaLeuProAlaValAlaGlnLysAlaAla 288

Qy   1219 TTGGGCTACCAGCTGCTCCGCCAGACCGCTACCAGTGCCGCTCCCGGAAGTCAATATTACT 1278
          :::          |||                       :::
Db    289 AsnAlaValSerMetThrValAsnTyrAsnProSerTrpIleGluGluAspHisHisPheThrVal 308

Qy   1279 GCCCCAGTTACTTCTTATTGGCCAACACAGACACCCGGATTCTTACAAACCTACCACCACTGGC 1338
          |||                                    :::||||
Db    309 ThrGluCysTyrPheProThrHisLysValGluGluAspHisPheThrVal------ 325

Qy   1339 ATCTACAGGTTCTACCCCATCTACATCAGCCGTCAGAAGCACTTCAAAACGCCATCAATTGGAGTC 1398
          ::::||||      :::
Db    326 -----------------------------------------------------LysAlaLeuLeuValAlaAsnAlaTyrArgLeu 333

Qy   1399 AACGAGAAAATCTCAGTCCAAGCCTATGAG 1428
          |||                  |||||||
Db                  LysAlaLeuLeuValAsnAlaTyrGlu 333
```

```
RESULT 11
ARGE_BUCAI

ID   ARGE_BUCAI     STANDARD;     PRT;    381 AA.
AC   P57155;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetylornithine deacetylase (EC 3.5.1.16) (Acetylornithinase) (AO)
DE   (N-acetylornithinase) (NAO).
GN   ARGE OR BU047.
OS   Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS   symbiotic bacterium).
OC   Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX   NCBI_TaxID=118099;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20445173; PubMed=10993077;
RA   Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT   "Genome sequence of the endocellular bacterial symbiont of aphids
RT   Buchnera sp. APS.";
RL   Nature 407:81-86(2000).
CC   -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + H(2)O = acetate + L-
CC       ornithine.
CC   -!- COFACTOR: COBALT AND GLUTATHIONE (BY SIMILARITY).
CC   -!- PATHWAY: Arginine biosynthesis; fifth step.
CC   -!- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
```

```
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; AP001118; BAB12770.1; -.
DR   MEROPS; M20.UNP; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; FALSE_NEG.
KW   Arginine biosynthesis; Cobalt; Hydrolase.
SQ   SEQUENCE   381 AA;  43045 MW;  A837B49131A1E0ED CRC64;
```

Alignment Scores:

```
Pred. No.:          0.00502
Score:              143.50
Percent Similarity: 38.04%
Best Local Similarity: 21.78%
Query Match:        5.20%
DB:                 1
```

US-10-014-896-1 (1-1509) x ARGE_BUCAI (1-381)

```
                                                   Length:        381
                                                   Matches:        71
                                                   Conservative:   53
                                                   Mismatches:    135
                                                   Indels:         67
                                                   Gaps:           12
```

```
Qy   157 GAGGCGCTGAAAGGTGCCATCCAGATTCCAACACAGTGACTTTTAGCTCT----------GAG 207
         ||||                                                        |||
Db    10 GluValThrPheLysSerLeuIleGlnIlePrOThrThrLeuSerSerAsnAsnLysLeuLeuAsp  29

Qy   208 AAGTCCAAATACTACAGCCCCTGGCTGAGTTCGGAAAAATACATTCATAAAGTCTTTCCTACA 267
         :::                                     |||
Db    30 GlnSerAsnLysAsnPheIleLeuSerAsnTyrPheSerAspLeuAsnPheSer  49

Qy   268 GTGGTCAGCACCAGCTTTATCCAGCATGAAGTGCGTGGAAGAAGATARGCCACCTGTTCACT 327
         |||:::      :::        :::||||:|:|          |||
Db    50 ValLysAsnTyrGln---IleProHis-----ThrAspThrTyrAsnMetLeuAlaCys  66

Qy   328 ATCCAAGGCTCCGACCCCAGCTTGCACCCTACCTCGATGGCTCCACTTTGATGTGTGGTG 387
         ||||||      :::|||    |||
Db    67 ValGlySerGlyAsnGlyGlyLeu-----LeuLeuSerGlyHisSerAspThrVal  83

Qy   388 CCTGCCCCTGAAGAAGGCCTGGGAGGTGCCCCCATTCTTCTGGGTTGGAGCGTGATGGCGTC 447
         |||:::       ||||||                 :::   |||
Db    84 AspPheAspGluLysLysTrpThrTyrAspProPheLysLeuThrAsnAsnLys 103

Qy   448 ATCTATGGTCGGGGACACTGGACGACAAGAAACTCTGTGATGGCAATTACTGCCAGGCCTTG 507
         ||||||      :::|||                |||:::       :::|||
Db   104 PheTyrGlyLeuGlyAlaValAlaAspMetLysGlyPhePheAlaLeuIleLeuGluValIle 123

Qy   508 GAGCTCTGCTCATCAGGAAGTACAATCCCCCGAAGATCTTTCTTCATTCTCTGGGCCAT 567
         :::       ||||||                |||
Db   124 SerSerIleAsnIleLysLysIle----------AsnLysProIleTyrIleLeuLeuAlaThrArgAla 141

Qy   568 GATGAGGAGTCATCAGGGACGAGGGGTTCAGGAGGATCTTCAGCGCCTGCTTACAGTCAAGGGGC 627
         :::|||:||:||||||                |||:::       |||:::
Db   142 AsnGluGlUThrAspMetSerGlyAlaArgAsn------PheIleGlnSerThrIle 158

Qy   628 GTCCAGCTAGCCTTCATTGTGGACGAGGAGGGGGCTTCATCTGGATGATTCATTCCTAAC 687
         |||:::||:       |||                |||:::
Db   159 IleLysProAspCysIleIle-----------TcAGAGAAGGGTTCCATGAACCTCATG 165

Qy   688 TTCAAGAAGCCCAATGCCCTTGATTGCAGTC---TCAGAGAAGGGTTCCATGAACCTCATG 744
         :::||||||      :::|||       ||| |||:::
Db   166 IleGlyGluProThrSerLeuLysLeuIleAsnAlaHisLysGlyHisMetSerTyrSer 185

Qy   745 CTGCAAGTAAAACATGACTTCAGCGCCCACTCTTCAGCTCCTCCAAAG---GAGACAAGCATT 801
         :::|||||||      :::|||                |||
Db   186 IleLysValIleGlyHisSerSerAsnProAspHisGlyValAsnSerIle 205

Qy   802 GGCATCCTTGCAGCTGCTGTCAGCCGGATTGGGAGCAGAGAGACCAACAAATGCCTATCATATTTGGA 861
         |||:::                |||                |||
Db   206 GluIleMetHisAspValIleArgSerLeu--------------------------------- 215
```

RESULT 12
ARGE_ECOLI

```
ID   ARGE_ECOLI    STANDARD;    PRT;   383 AA.
AC   P23908;
DT   01-MAR-1992 (Rel. 21, Created)
DT   01-DEC-1992 (Rel. 24, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Acetylornithine deacetylase (EC 3.5.1.16) (Acetylornithinase) (AO)
DE   (N-acetylornithinase) (NAO).
GN   ARGE OR B3957.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92202162; PubMed=1551850;
RA   Meinnel T., Schmitt E., Mechulam Y., Blanquet S.;
RT   "Structural and biochemical characterization of the Escherichia coli
RT   argE gene product.";
RL   J. Bacteriol. 174:2323-2331(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=92331940; PubMed=1628835;
RA   Boyen A., Charlier D.R.M., Sakanyan V., Mett I., Glansdorff N.;
RT   "Acetylornithine deacetylase, succinyldiaminopimelate desuccinylase
RT   and carboxypeptidase G2 are evolutionarily related.";
RL   Gene 116:1-6(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=K12 / MG1655;
RX   MEDLINE=94089392; PubMed=8265357;
RA   Blattner F.R., Burland V.D., Plunkett G. III, Sofia H.J.,
RA   Daniels D.L.;
RT   "Analysis of the Escherichia coli genome. IV. DNA sequence of the
RT   region from 89.2 to 92.8 minutes.";
RL   Nucleic Acids Res. 21:5408-5417(1993).
RN   [4]
RP   SEQUENCE OF 1-11 FROM N.A.
RC   STRAIN=K12;
RX   MEDLINE=83143275; PubMed=6761650;
RA   Piette J., Cunin R., Boyen A., Charlier D.R.M., Crabeel M.,
RA   van Vliet F., Glansdorff N., Squires C., Squires C.L.;
RT   "The regulatory region of the divergent argECBH operon in Escherichia
RT   coli K-12.";
RL   Nucleic Acids Res. 10:8031-8048(1982).
RN   [5]
RP   SEQUENCE OF 1-11 FROM N.A.
RX   MEDLINE=83064529; PubMed=6292860;
RA   Charlier D.R.M., Piette J., Glansdorff N.;
RT   "IS3 can function as a mobile promoter in E. coli.";
```

```
Qy   862 AGCGGGACAGTGGTGACTGTGATTGGCAGCAACTGGCAAATGAGTTTCCCTTTGCTTGTCAAT 921
         :::       :::           |||:::          |||
Db   216 ----------LeuIleLeuLysTyrPheLysGluGlyTyrGlnHisPro--Asn 230

Qy   922 ATAATCCTGAGCAACCCATGGCTATTGACCAACTTATAAGCAGGTTTATGGAGAGAAAT 981
         :::         ||||||:::                |||
Db   231 PheSerIleProTyrPro--------------------------------- 236

Qy   982 CCCTTAACCAATGCAATAATCAGGACCACACGGCGCTCACCATATTCAAAGCAGGGGTC 1041
         |||                |||:::          :::|||
Db   237 ----------ThrMetAsnLeuSerSerIleHisGlyGlySer 247

Qy  1042 AAGTTCAAATGTCAATCCCCCAGTGGCCCCAGGCCCACAGTCAACTTCCGGATTCACCTGGA 1101
         |||:::       |||:::          |||               |||
Db   248 AlaIleAsnArgIleCysProLeuCysIleLeuAsnPheGluIleLeuArgProIleProGly 267

Qy  1102 CAGACAGTCCAAGAGGTC 1119
         |||:::|||:::
Db   268 LeuThrLeuThrGlnIle 273
```

```
RL   Nucleic Acids Res. 10:5935-5948(1982).
CC   -!- FUNCTION: DISPLAYS A BROAD SPECIFICITY AND CAN ALSO DEACYLATES
CC       SUBSTRATES SUCH AS ACETYLARGININE, ACETYLHISTIDINE, OR
CC   -!- ACETYLGLUTAMATE SEMIALDEHYDE.
CC       ornithine.
CC   -!- CATALYTIC ACTIVITY: N(2)-acetyl-L-ornithine + H(2)O = acetate + L-
CC       ornithine.
CC   -!- COFACTOR: COBALT AND GLUTATHIONE.
CC   -!- PATHWAY: Arginine biosynthesis; fifth step.
CC   -!- SUBUNIT: HOMODIMER.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20B.
CC   -----
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----
DR   EMBL; X55417; CAA39076.1; -.
DR   EMBL; X62807; CAA44625.1; -.
DR   EMBL; U00006; AAC43063.1; -.
DR   EMBL; AE000470; AAC76939.1; -.
DR   EMBL; J01587; AAB59145.1; -.
DR   EMBL; J01589; AAA23484.2; -.
DR   PIR; B42377; B42377.
DR   PIR; S23278; S23278.
DR   MEROPS; M20.UNB; -.
DR   EcoGene; EG11286; argE.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; 1.
KW   Arginine biosynthesis; Hydrolase; Cobalt; Complete proteome.
SQ   SEQUENCE   383 AA;  42347 MW;  AFE04B99B296540B CRC64;
```

Alignment Scores:

```
Pred. No.:
Score:                 0.0118
Score:                 138.50
Percent Similarity:    33.42%
Local Similarity:      22.91%
Best Local Similarity: 5.01%
Query Match:           1

Length:         383
Matches:        85
Conservative:   39
Mismatches:     140
Indels:         108
Gaps:           15
```

```
us-10-014-896-1 (1-1509) x ARGE_ECOLI (1-383)

Qy   361 CTGCTGATGGCTCACTTTGATGGTGGTCGCCTGCCCTGAAGAAGGCTGGGAGGTGCCCCCA 420
             |||||| |||   |||  ||||||||| |||  |||||||||||||||    |||
Db    76 LeuLeuAlaGlyHisThrAspThrValProPheAspAspGlyArgTrpThrArgAspPro  95

Qy   421 TTCTCTGGGTTGGAGCGTGATGGCCGTCATCTATGGTCGGGGGCACACTGGACGACAAGAAC 480
Db    96 PheThrLeuThrGluHisAspSerHisTyrGlyLeuGlyThrLeuAlaAspMetLysGly 115

Qy   481 TCTGTGATGGCATTACTGCGAGGCCTCCTGATCAGGAAGTACATACCCCGA 540
Db   116 PhePheAlaPheIleLeuAspAlaLeuArgAspValAspValThrGlyLys 133

Qy   541 AGATCTTTCTTCATTCTCTGGGCCCATGAGAGAGTCATCAGGGACAGAGGGCTCAGAGG 600
Db   134 LysProLeuTyrIleLeuAlaThrAlaAspGluGluThrSerMetAlaGlyAlaArgTyr 153

Qy   601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTTGACGAGGGGGGC 660
Db   154 PheAlaGluThrThrAlaLeuArg------ProAspCysAlaIleIleIleGlyGlu--- 169

Qy   661 TTCACCTTGATGATTCAATTCCTAACTCAAGAAGCCCATCGCCTTGATTGCAGTCTCA 720
Db   170 ------ProThrSerLeuGlnProValArg-------Ala 178

Qy   721 GAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCCACTCTTCAGCT 780
          ||||||                                              ::::::
Db   --- ----------------------------------------------------------

Db   179 HisLysGlyHisIleSerAsnAlaIleArgIleGlnGlyGlnSerGlyHisSerSerAsp 198
             :::
Qy   781 CCTCCAAAGGAGGACAAGCCATTGGCAGTCCAGCCGTTGCAGCTGCGTCAGCCGATTGGAGAGACA 840
                                                                 ||
Db   199 Pro-------AlaArgGlyValAsnAlaIleGluLeuMet 209

Qy   841 CCAATGCCTATCATATTTGGAAGCGGGACAGTGGGACTGTATTGCGACAACTGGCAAAT 900
Db   210 HisAspAlaIle-------GlyHisIleLeuLeuArgAspAsnLeuLysGlu 225

Qy   901 GAGTTT------CCCTTCCCTGGTCAAATAATACCTGAGCAACCCA 939
             ||| :||:
Db   226 -------ThrLeuAsnLeuGlyHisIleHisGlyLysCys 237

Qy   940 TGGCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAGAAATCCCTTACCAATGCAATA 999
Db   237 ------------------------------------------------- 237

Qy  1000 ATCAGGACCACACGGCCACTCACCATATTCAAAAGCAGGGGTGGTCAAGTTCAATGTGTCAATCCCC 1059
             |||
Db  1000 -------------------------------------------------------------------

Db  238 -------ThrLeuAsnLeuGlyHisIleHisGlyLysCys 254

Qy  1060 CCAGTGCGCCCAGGCCACAGTCCAACTTCCGGAATTCACCCTGGACAGACAGTCCAAGAGGTC 1119
             ||||:::||          |||:::||
Db  254 ---CAGAGTCCAGTTCCATTGTGTTAGTGCCTTGAGCCCCTCCCGTCAG 1199

Qy  1120 CTAGAACTCAGGAAGAACACAATTGTGGCTGATAA 1151
             |||
Db  1120 -----------------------------------

Db  255 AlaCysCysGlyLeuHisMetAspIleArgProLeuProGlyMetThrLeuAsnGluLeu 274

Qy  1152 CAGAGTCCAGTTCCATTGTGTTAGTGCCTTGAGCCCCTCCCGTCAG 1199
Db  1152 --------------------------------------------

Qy  1200 ------CCCTTCTGATGACAAGGGCCTTGGGCTACCAGCGTGCTCCGGCACACGCGTACAGTC 1253
             |||||
Db  1200 GlnAlaHisGlnProAspGluTyrLeuGluThrArgPheIleLysProThrArgGluLeu 372

Qy  1254 CGTCGTTCCCGGAAGTCAAAATTACTGCGCCCAGTTACTTCTATTGG 1298
Db  1254 --------------------------------------------

Db  275 AsnGlyValLeuLeuAsnAspSerLeuLeuAlaProValSerGlyArgTrpProGlyLeuThr 294

Qy  1299 CAACACAGACAGCCGGATTCTTTAC 1328
             ||||||           ||
Db  1299 --------------------AAACCT------ 1328

Db  295 ValAspGluHisGlyProThrIleProGlyValTyrArgGlyCys------ProProAsnHisGln 312

Qy  1329 ------CACCAC-----TGGCAT 1340
             |||||||  ::
Db  1329 ----------------------

Db  313 LeuValGluValValGlyLysLeuLeuGlyLyrAlaLysThrGluValValAlaAsnTyrCysThr 332

Qy  1152 CAGAGTCCAGTTCCATTGTGTTAGTGCCTTGAGCCCCTCCCGTCAG 1199
Db  333 GluAlaProPheIleGlnThrLeuCysProThrTrpLeuValLeuGlyProGlySerIleAsn 352

Db  353 GlnAlaHisGlnProAspGluTyrLeuGluThrArgPheIleLysProThrArgGluLeu 372

Db  373 IleThrGlnIleValIleHisHisPheCysTrpHis 383
```

RESULT 13
YFL4_YEAST

```
ID   YFL4_YEAST     STANDARD;    PRT;   481 AA.
AC   P43616;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-NOV-1995 (Rel. 32, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical 52.9 kDa protein in SAP155-YMR31 intergenic region.
GN   YFR044C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=95400292; PubMed=7670463;
RA   Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA   Sasanuma S.-I., Sasanuma M., Tsuchiya Y., Soeda E., Yokoyama K.,
RA   Yamazaki M., Tashiro H., Eki T.;
RT   "Analysis of the nucleotide sequence of chromosome VI from
RT   Saccharomyces cerevisiae.";
RL   Nat. Genet. 10:261-268(1995).
```

```
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288C / AB972;
RX   MEDLINE=96287652; PubMed=8886379;
RA   Eki T., Naitou M., Hagiwara H., Ozawa M., Sasanuma S.-I.,
RA   Sasanuma M., Tsuchiya Y., Shibata T., Hanaoka F., Murakami Y.;
RT   "Analysis of a 36.2 kb DNA sequence including the right telomere of
RT   chromosome VI from Saccharomyces cerevisiae.";
RL   Yeast 12:149-167(1996).
CC   -!- SIMILARITY: STRONG, TO THE C-TERMINAL HALF OF YEAST YBR281C.
```

```
DR   EMBL; D50617; BAA09283.1; -.
DR   SGD; S0001940; YFR004C.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   481 AA;  52871 MW;  3E53773A945F5EBC CRC64;

US-10-014-896-1 (1-1509) x YFI4_YEAST (1-481)

Alignment Scores:
Pred. No.:                                Length:        481
Score:              0.025                 Matches:        54
                  134.50                  Conservative:   40
Percent Similarity: 41.05%                Mismatches:     66
Best Local Similarity: 23.58%             Indels:         69
Query Match:        4.87%                 Gaps:           10
                    1
```

```
Qy  118 CCTTCTCAGTTCAGCAAAGAGGAACGGCGTCGGATGAAAAGAGGCGCGCTGAAAGGTGCCCATC 177
            |||  ||||||||  ||::::
Db   17 ProGlnPhePheSerArg----------------------LeuThrLysAlaIle 27

Qy  178 CAGATTCCACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACTACAGCCCTGGCTGAGTTC 237
            |||||||||  ||    ::  :::::::::||
Db   28 GlnIleProAlaVal----SerSerAspGluSerLeuArgSerLysValPheAspLys 45

Qy  238 GGAAAATACATTCATAAA---------------------- 255
            ||:::|||::|||||||
Db   46 AlaLysPheIleSerGluGlnLeuSerGlnSerGlyPheHisAspIleLysMetValAsp 65

Qy  256 ------GTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAG 309
                  :::  ||||||:|
Db   66 LeuGlyIleGlnProProIleSerThrProAsnLeuSerLeuProProVal----- 83

Qy  310 TATAGCCACCTGTTCACTATCCAAGGCTCGGAGCCCAGCGTTGCAGCCCTACCTGCTGATG 369
            |||||||||       |||      ::
Db   84 ---------IleLeuSerArgPheGlySerAspProSerLysLysThrValLeuValTyr 100

Qy  370 GCTCACTTTGATGTGTGCCCCCT----GAAGAAGGCGTGGGAGGTGCCCCATTCTCT 426
            ||:::||:::|||||||  |||:::||||||||||:::|||:::    |||||||
Db  101 GlyHisTyrAspValGlnProLeuGlnLeuGluAspGlyTrpAspThrGluProPheLys 120

Qy  427 GGGTG------GAGCGTGATGGCGTGCATCTATGGTTCGGGCCACACTGGACGACAAGAAC 480
            |||      :::|||:::||||||  :::||||||||
Db  121 LeuValIleAspGluAlaLysGlyIleMetLysGlyAspIleArgGlyValThrArgAspThrGly 140

Qy  481 TCTGTGAATGGGCAATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGGAGGTACATCCCCGA 540
            ||  :::  ||
Db  141 ProLeuLeuSerTrpIleAsnValValAspPro----- 151

Qy  541 AGATCTTTTCTTCATTTCTCTGGGCCATGAT--------- 573
            |||       |||                      :::GAG 573
Db  152 ------PheLysAlaSerGlyGlnGluPheProValAsnLeuValThrCysPheGlu 168

Qy  574 GAGTCATCAGGGACACAGGGCTCAGAGGATCTCAGCCCTGCTACAGTCA----- 621
```

```
Db  169 GlyMetGluGluSerGlySerLeuLysLeuAspGluLeuIleLeuLysGlyAlaAsnGly 188
            :::||::::|    ::::::    ::::::
Qy  622 -------AGGGGCGTCCAGCTAGCCTTCATTGTGGACGAGGAGGGGGCTTCATCTTGGATGAT 675

Db  189 TyrPheLysGlyVal------------------------AspAlaValCysIleSerAspAsn 201
            |||:::                                         |||:::
Qy  676 TTCATTCCTAAACTTCAAGAAGCCCATC 702

Db  202 TyrTrpLeuGlyThrLysLysProVal 210
```

```
RESULT 14
YS89_CAEEL

ID   YS89_CAEEL      STANDARD;      PRT;    3178 AA.
AC   Q09624; Q09625; Q069D4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   15-JUN-2002 (Rel. 41, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hypothetical protein ZK945.9 in chromosome II.
GN   ZK945.9/ZK945.10.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Wilkinson-Sproat J.;
RL   Submitted (FEB-1995) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   REVISIONS.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC   -!- SIMILARITY: CONTAINS 1 GPS DOMAIN.
```

```
DR   EMBL; Z48544; CAB70192.1; -.
DR   EMBL; Z48582; CAB70192.1; -.
DR   EMBL; Z48582; CAB70201.1; JOINED.
DR   EMBL; Z48544; CAB70201.1; -.
DR   EMBL; Z48544; CAB70201.1; JOINED.
WormPep; ZK945.9; CE25697.
DR   InterPro; IPR002111; Cat_channel_TrpL.
DR   InterPro; IPR001024; Lipoxygenase_LH2.
DR   InterPro; IPR000636; M+channel_nlg.
DR   InterPro; IPR000203; PKD_cys_rich.
DR   Pfam; PF00520; Ion_trans; 1.
DR   Pfam; PF01477; PLAT; 1.
DR   Pfam; PF01825; GPS; 1.
DR   SMART; SM00303; GPS; 1.
DR   SMART; SM00308; LH2; 1.
KW   Hypothetical protein; Transmembrane.
FT   DOMAIN      266   1196       SER/THR-RICH.
FT   DOMAIN     1105   1241       GLY/SER-RICH.
FT   DOMAIN     2071   2120       GPS.
FT   TRANSMEM     51     73       POTENTIAL.
FT   TRANSMEM   2139   2161       POTENTIAL.
FT   TRANSMEM   2348   2367       POTENTIAL.
FT   TRANSMEM   2390   2412       POTENTIAL.
FT   TRANSMEM   2451   2468       POTENTIAL.
FT   TRANSMEM   2483   2505       POTENTIAL.
FT   TRANSMEM   2567   2589       POTENTIAL.
FT   TRANSMEM   2836   2858       POTENTIAL.
FT   TRANSMEM   2939   2961       POTENTIAL.
FT   TRANSMEM   2976   2998       POTENTIAL.
```

```
FT   TRANSMEM      3038   3060
SQ   SEQUENCE      3178 AA;  344726 MW;  F8239436D03666CD CRC64;

US-10-014-896-1 (1-1509) x YS89_CAEEL (1-3178)

Alignment Scores:
  Pred. No.:                     1.01
  Score:                       116.00
  Percent Similarity:          38.31%
  Percent Identity:            24.10%
  Best Local Similarity:        4.20%
  Query Match:                      1

  Length:                        3178
  Matches:                        100
  Conservative:                    59
  Mismatches:                     188
  Indels:                          69
  Gaps:                            14
```

```
Qy         1 GCCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTCTGGGTTGGAGCGT---GATGGCGTC 447
             |||||  ||                ||||::|||||||          |||  |||||||
Db       391 GACTCCTGCTCAATCAGGAAGTACACATCCCCGAAAGATCTTTCTTCATTTCTCTGGGCCAT 567

Qy       448 ATCTATGGTGTCGGGGCGGCCACACTCGATGACGACAAGAACTCTGTGGCATTACTCCTAAC 507
             :::::                               |||::                 |||
Db       179 Leu Trp ------         -----Trp Leu Gln Ile Val Gly Ile Gly Ile Ala Asn Glu Met Ala Ile Leu Leu Gly Thr Tyr Arg 195
```

```
RESULT 15
HIPO_CAMJE
ID   HIPO_CAMJE      STANDARD;      PRT;    383 AA.
AC   P45493; Q9PNV4;
DT   01-NOV-1995 (Rel. 32, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-JUN-2002 (Rel. 41, Last annotation update)
DE   Hippurate hydrolase (EC 3.5.1.32) (Benzoylglycine amidohydrolase)
DE   (Hippuricase).
GN   HIPO OR CJ0985c.
OS   Campylobacter jejuni.
OC   Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC   Campylobacter.
OX   NCBI_TaxID=197;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 43431 / TGH 9011;
RX   MEDLINE=95247673; PubMed=7730270;
RA   Hani E.K., Chan V.L.;
RT   "Expression and characterization of Campylobacter jejuni
RT   benzoylglycine amidohydrolase (Hippuricase) gene in Escherichia
RT   coli.";
RL   J. Bacteriol. 177:2396-2402(1995).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NCTC 11168;
RX   MEDLINE=20150912; PubMed=10688204;
RA   Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA   Basham D., Chillingworth T., Davies R.M., Feltwell T., Holroyd S.,
RA   Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA   Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA   Whitehead S., Barrell B.G.;
RT   "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT   reveals hypervariable sequences.";
RL   Nature 403:665-668(2000).
CC   -!- CATALYTIC ACTIVITY: Hippurate + H(2)O = benzoate + glycine.
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M40.
DR   EMBL; Z36940; CAA85396.1; -.
DR   EMBL; AL139076; CAB73241.1; -.
DR   MEROPS; M40.UNW; -.
```

DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
KW   Hydrolase; Complete proteome.
FT   CONFLICT   125   125   A -> T (IN REF. 1).
FT   CONFLICT   213   213   A -> V (IN REF. 1).
FT   CONFLICT   250   250   A -> V (IN REF. 1).
FT   CONFLICT   278   278   I -> V (IN REF. 1).
FT   CONFLICT   278   278   I -> L (IN REF. 1).
SQ   SEQUENCE   383 AA; 42546 MW; 01C94C1293F6B1C8 CRC64;

Alignment Scores:
Pred. No.:   0.605
Score: 115.50
Percent Similarity: 37.14%
Best Local Similarity: 18.45%
Query Match: 4.18%

Length: 383
Matches: 76
Conservative: 77
Mismatches: 130
Indels: 129
Gaps: 20

US-10-014-896-1 (1-1509) x HIPO_CAMJE (1-383)

```
Qy   397 GAAGAAGGCTGGGAGGTGCCCCAATTCTCTGGGTTGGACGGTGATGGCGGTC------ATC 450
         ||| ||| :||||| |||||||  ::  ::  |||| :::  |||||:|  |||    |||
Db    44 GluPheGlyTyrGluVal------TyrGluGluIleGlyLysThrGlyValValGlyVal  61

Qy   451 TATGGTTCGGGGGCACACTGACGACAAGAACTCTGACTGGCATTACTGCAGGCCTTG--- 507
         :::|||| ||  ||| |:||||  | |||  || ||||  ::|  ||| ::: |||
Db    62 LeuLysLysGlyAsnSerAspLysLysIleGlyLeuArgAlaAspMetAspAlaLeuPro  81

Qy   507 ------------------------------------------------------------ 507
Db    82 LeuGlnGluCysThrAsnLeuProTyrLysSerLysLysGluAsnValMetHisAlaCys 101

Qy   508 -----------GAGCTCCTGCTGATCAGGAGGAAGTACAATCCCCCGAGAGATCT 546
                    |||||||| ::  ::  |||  ||| ::::::
Db   102 GlyHisAspProGlyHisThrThrSerLeuLeuLeuAlaLysTyrLeuAlaSerGlnAsn 121

Qy   547 TTC----TTCATTTCTCTGGGCCATGATGAGGAGTCATCAGACAGAAGGTTCCATGAAC 600
         |||    ::::|||||| ::|||  || ::   |:: ||| ||  |||| |||:
Db   122 PheAsnGlyAlaLeuAsnLeuTyrPheGlnProAlaGluGlyLeuGlyLysAlaLys 141

Qy   601 ATCTCAGCGCCTGCTACAGTCAGAAGGGGCGTCCAGCCTTCAGTCTGGACGAGGGGGGC 660
         ::::::::::|||
Db   142 Ala---------------------------MetIleGluAspGlyIleu 148

Qy   661 TTC-----AATCTTGGATGATTTCATT-------------------------------- 681
         |||     :::|||  ||  |:||
Db   148 MetIleGluAspGlyIleu 148

Qy   682 ---CCTAAACTTCAAAGAACGCCATCCGCCTTGAATTGCAGTCTCAGAGAAAGGTTCCATGAAC 738
            ::|  |||||| ::|||  || ::  |::   |||  | :::  || || ||| :::
Db   169 LysLysPheTyrLeuLysGlyAlaMetHisAlaSerSerAspSerTyrSer 186

Qy   739 CTCATGCTGCAAGTAAACACNGACTTCAGGCCACTCTTCAGCTCCTCCAAAGGAGACAAGC 798
         ||||||  :::::::::|||  ||| ||||| ||| || ::| |||
Db   187 -----IleGluValIleGlyAlaArgGlyIleGlyHisGlySerAlaProGluLys 201

Qy   799 AATGGGACATCCTTGCAGCTGCTGCTGCTCAGCGGATTGGAGCGACACCAATGCCTATCATATTT 858
          ::    :::   ::|    ::  :::  || ||  |||  || ||| ::  |||
Db   202 ---------------------AlaLysAspPro----IleTyr 207

Qy   859 GGAAGCGGGACAGTGTTGACTGTATTGCAGCAAATGAGTTTTCCCTTCCCTGTC 918
         |||||  |  :::|| ||| :::  ||  :::  || ::|  |||::::
Db   208 AlaAlaSerLeuLeuIleLeuValAlaLeuGlnSer 218

Qy   919 AATATAATCCTGAGCAACCACNNCCATGGCTATTTTGAACACNTTATNAAGCAGGTTTATTGGAGAGA 978
         ::::::::::|||:::  || :: |||  ||  |||  :::|| ||||  |||  ||| |||
Db   219 -----IleLeuSerArgAsnValAsp--- 225

Qy   979 AATCCCTTAACCAAATGCAATAATCAGGACGACCACACNGGCACTCACCATANTATTCAAAGCAGGG 1038
         |||  |  |||  :::  ||  || ::| ||| ||| :: ::: ::| :::  || :::  |||
Db   226 ---ProGlnAsnSerAlaValVal 240

Qy  1039 GTCAAGTTCAAATGTCATCCCCCAGTGGGCCCACAGTCAACTTCCGGATTCACCCT 1098
         ||||:::   :::|||  || ::|||  ||| ||  ||| || |||  |:||
Db   241 HisAlaPheAsnIleIleProAspIleAlaThrIleLeuLysMetSerValArgAlaLeuAsp 260

Qy  1099 GGACAG------ACAGTCCAAGAAGGTCCTTAGAACTCACGAGAACATTGTGGCT 1146
         ::::|||||: |||||| ::::||  |||||::| |||  ::  :::  ||||||
Db   261 AsnGluThrArgLysLeuThrGluGluLysIleCysLysGlyIleLeuAlaGln 280

Qy  1147 GATAACAGAGTCCAGTTCCATGTGTTGAGGCGCTTGACCCCCTCCCGGCCCTTCT 1206
         ||| |||||  |||  || ::  :::   ||  || ::|||  ::::::  :::
Db   281 AlaAsnAspIleGluIleLysIle---AsnLysAsnValValAlaProValThrMetAsn 299

Qy  1207 GATGACAAGGCCTTGGGCTACCAGCTGCTCCGGCAGACCTGCTACAGTCCGGTTCTTCCGGAA 1266
         ||| ::||||  :::|||  || || ::   ||  ||  ::  || :::  || ::|
Db   300 AsnAspGluAlaValAlaAspPhe------AlaSerGluValAlaLysGluLeuPheGlyGlu 317

Qy  1267 GTCAATATTACT--------GCCCCAGTTACTCTCTTATTGGGCACACAACAGCAGCCGGATTC 1317
         ::::::::::         ||||  ||  ||| || :::   |||  ||   ||  || |||
Db   318 LysAsnCysGluPheAsnHisArgProLeu-----MetAlaSerGluAspProGlyPhe 335

Qy  1318 TTTACAAAACCTCACCACTGCCATCTACAGGTTCTCGACCCCATCACATACAGAGCCTGAAGAC 1377
         |||  :::||| ||| :::||| ::|||  ||  || ||| ::  :::   :::|||
Db   336 PheCysGluMetLysLysCysAlaTyrArgAlaPhe------LeuGluAsn 351

Qy  1378 TTCAAACGCATCCATGGA--------GTCACGAGAAAATCTCAGTCCAAGCCTAT 1425
         ||| ::|  ::||| |||         :::  ::|||  |  | ::|||  :::
Db   352 AspIleTyrLeuHisAsnSerTyrValPheAsnAspLysLeuLeuAlaArgAlaAla 371

Qy  1426 GAGACCCAAGTGAAAATCATCTTTGAGTTGATTCAG 1461
         ||| ::| :::   ::||| || |||  |::||
Db   372 SerTyrTyrTyrAlaAlaGlnLeuAlaLeuValLysTyrLeuLys 383
```

Search completed: June 27, 2003, 11:03:22
Job time : 49 secs

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 27, 2003, 10:46:38 ; Search time 125.5 Seconds
                                      (without alignments)
                                      4954.982 Million cell updates/sec

Title: US-10-014-896-1
Perfect score: 2762
Sequence: 1 atggtcagcggtgcgttg..............ctcacctgcacaaactgtga 1509

Scoring table: BLOSUM62
                Xgapop 10.0 , Xgapext   0.5
                Ygapop 10.0 , Ygapext   0.5
                Fgapop   6.0 , Fgapext   7.0
                Delop   6.0 , Delext   7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104414_10381/app_query.fasta_1.1671
-DB=SPTREMBL_21 -QFMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=blosum62 -TRANS=human40.cd1
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10014896 @CGN_1_1_172_@runat_27062003_104414_10381 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :   SPTREMBL_21:*
            1   sp_archea:*
            2   sp_bacteria:*
            3   sp_fungi1:*
            4   sp_human:*
            5   sp_invertebrate:*
            6   sp_mammal:*
            7   sp_mhc:*
            8   sp_organelle:*
            9   sp_phage:*
           10   sp_plant:*
           11   sp_rodent:*
           12   sp_virus:*
           13   sp_vertebrate:*
           14   sp_unclassified:*
           15   sp_rrvirus:*
           16   sp_bacteriap:*
           17   sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1816	65.7	361	4	Q96DM4	Q96dm4 homo sapien

2	1261.5	45.7	340	16	Q8R117	Q8r117 mus musculu
3	781	28.3	510	16	Q8XS17	Q8xs17 ralstonia s
4	639.5	23.2	488	2	Q8RTT1	Q8rtt1 uncultured
5	591	21.4	471	16	Q9AA70	Q9aa70 caulobacter
6	531.5	19.2	596	3	O13968	O13968 schizosacch
7	434	15.7	469	2	Q8RNM5	Q8rnm5 legionella
8	361.5	13.1	416	16	Q92BI7	Q92bi7 streptomyce
9	341.5	12.4	474	16	Q9A3G5	Q9a3g5 caulobacter
10	330.5	12.0	467	16	Q9CC46	Q9cc46 mycobacteri
11	318	11.5	441	2	Q93H22	Q93h22 streptomyce
12	309.5	11.2	448	16	O06234	O06234 mycobacteri
13	300	10.9	445	16	Q9AD91	Q9ad91 streptomyce
14	276	10.0	442	16	Q93R29	Q93r29 streptomyce
15	253.5	9.2	401	5	Q9VCQ9	Q9vcq9 drosophila
16	232.5	8.4	382	17	Q97OS3	Q97os3 sulfolobus
17	229.5	8.3	397	5	Q17899	Q17899 caenorhabdi
18	228	8.3	400	5	Q9VCR0	Q9vcr0 drosophila
19	227.5	8.2	402	5	O8T490	O8t490 drosophila
20	225.5	8.2	378	16	Q92FI9	Q92fi9 listeria in
21	224.5	8.1	408	5	Q9VCR1	Q9vcr1 drosophila
22	223	8.1	410	17	Q97YI2	Q97yi2 sulfolobus
23	217	7.9	340	5	Q9VCQ8	Q9vcq8 drosophila
24	214.5	7.8	382	16	Q92FY0	Q92fy0 rickettsia
25	213.5	7.7	455	17	O59016	O59016 pyrococcus
26	212	7.7	438	10	Q9LPE9	Q9lpe9 arabidopsis
27	209	7.6	435	10	Q9C6Y8	Q9c6y8 arabidopsis
28	207.5	7.5	359	5	Q9VCR2	Q9vcr2 drosophila
29	207.5	7.5	359	5	Q17898	Q17898 caenorhabdi
30	205.5	7.4	397	5	O8XWW5	O8xww5 ralstonia s
31	205	7.4	383	16	Q9IH5	Q9ih5 pseudomonas
32	204	7.4	408	16	Q9CRI5	Q9cri5 mus musculu
33	203	7.3	379	16	Q92EY0	Q92ey0 listeria mo
34	201.5	7.3	397	17	O82VD7	O82vd7 pyrobaculum
35	200.5	7.3	428	5	O17686	O17686 caenorhabdi
36	197.5	7.2	383	16	Q92C93	Q92c93 rickettsia
37	197.5	7.2	401	16	Q9I7K3	Q9i7k3 drosophila
38	197	7.1	374	16	Q92Y75	Q92y75 rhizobium
39	197	7.1	379	2	Q9EXF4	Q9exf4 listeria mo
40	196.5	7.1	394	5	O17900	O17900 caenorhabdi
41	196.5	7.1	407	16	Q99SN6	Q99sn6 staphyloco
42	196	7.1	377	16	Q9KQ52	Q9kq52 vibrio chol
43	195.5	7.1	377	16	Q9TT10	Q9tt10 streptococc
44	194	7.0	388	16	Q92MM0	Q92mm0 helicobacte
45	194	7.0	408	11	Q99JW2	Q99jw2 mus musculu

RESULT 1
Q96DM4

ID   Q96DM4
AC   Q96DM4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   CDNA FLJ32269 fis, clone SPLEN2000134, weakly similar to
DE   carboxypeptidase S precursor (EC 3.4.17.4).
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=SPLEEN;
RA   Kawakami B., Sugiyama A., Takemoto M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Ota T., Wakamatsu A., Ishii S., Yamamoto J.,
RA   Isono K., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K.,
RA   Yamashita H., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K.,
RA   Wagatsuma M., Murakawa K., Kanehori K., Takahashi-Fujii A., Oshima A.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK057131; BAB71368.1; -.

PRELIMINARY;   PRT;   361 AA.

ALIGNMENTS

DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
SQ   SEQUENCE   361 AA; 39560 MW; 367DECB0991F314C CRC64;

Alignment Scores:
Pred. No.:
Score: 2.04e-144
                       1816.00
Percent Similarity: 93.73%
Percent Identity: 93.73%
Best Local Similarity: 65.75%
Query Match: 4

Length: 361
Matches: 359
Conservative: 0
Mismatches: 2
Indels: 2
Gaps: 1

DB:

US-10-014-896-1 (1-1509) x Q96DM4 (1-361)

Qy   1 ATGGCTCAGTCACAAAGAGGAACGCGTCGCGATGGAGAAGACAGGCGCCTGAAAGGTGCCCATCCAG   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MetAlaGlnArgCysValCysValAlaMetGluLysLysGlyAlaLeuLysGly   20

Qy   61 ACGGTCTCCAGATCGATGGGCCCGAGGAGCGGGGGAGCATCAAAGGGCGGTCGCGGAATCCCT   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   21 ThrValSerArgSerMetGlyProArgSerGlyGluHisGlnArgAlaSerArgIlePro   40

Qy   121 TCTCAGTTCAGCAAAGAGGAAGAAGAAGACAGGTCGCGATGAGAAGACAGGCGCCTGAAAGGTGCCCATCCAG   180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   41 SerGlnPheSerLysGluGluGluGluArgValAlaMetLysLysGlyAlaLeuLysGlyGln   60

Qy   181 ATTCCAACAGTGACTTTTAGCTCTCTGAGAAGTCCCAATACTACAGCGCCTGGCTGGTTCGGA   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 IleProThrValThrPheSerSerLeuArgSerProAsnThrThrAlaLeuAlaLeuAlaGlyPheGly   80

Qy   241 AAATACAATTCAAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCACCATGAAGTC   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   81 LysTyrIleHisLysValPheProThrValValSerThrSerPheIleGlnHisGluVal   100

Qy   301 GTGGAAGAGTATAGGCCACCTGTTCACTATCCACAAGGCTCGGACCCCAGCTTGCAGCCCTAC   360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   101 ValGluGluTyrSerHisLeuPheThrIleGlnGlySerAspProSerLeuGlnProTyr   120

Qy   361 CTGCTGATGGGTCACTTTGATGTGGGTGCCCTGCCCCTGAAGAAAAGGTGTGGGAGGTGCCCCA   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 LeuLeuMetAlaHisPheAspValValProLeuProLeuLysLysArgCysGlyLeuProPro   140

Qy   421 TTCTCTCTGGGTTGGAGCGCGTGATGAGGCCGTCATCTATGGTCGGGGCACACTGGGACAAGAAC   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   141 PheSerGlyLeuGluArgAspGlyValIleTyrGlyTrpGlyThrLeuGlyAspLysAsn   160

Qy   481 TCTGTGATGGCATTACTGCGAGGCCCTTGGAGCCTCCTGCTGATCAGGAAGTACATCCCCGGA   540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   161 SerValMetAlaLeuLeuGlnAlaLeuGluLeuLeuLeuIleArgLysTyrIleProAsp   180

Qy   541 AGATCTTTCATTTTCTGGGCCATGAGTGAGGGAGTCACTCAGGGAGACAGGGGTCGAGAGG   600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ArgSerPhePheIleSerLeuGlyHisSerGluGlyValThrGlnGlyAspArg   200

Qy   601 AACTCACCCCTGCTACAGTCAAGGGCGTCCAGCTAGCCCTTCATTGTGGGACGAGGGGGGGGC   660
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   201 IleSerAlaLeuLeuGlnSerArgGlyValGlnLeuAlaPheIleValAlaGluGlyGly   220

Qy   661 TTCACTGTTGGATGATGATTTCATTCCTAACTTCAAGAAGCCCATCGCCTTGAATTGCAGTCTCA   720
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   221 PheIleLeuAspAspPheIleProAsnPheLysLysProIleAlaLeuAsnCysSer   240

Qy   721 GAGAAGGGTTCCATGAACCCTCCATGTTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCT   780
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GluLysGlySerMetAsnLeuMetLeuGlnValAsnMetThrSerGlyHisSerSerAla   260

Qy   781 CCTCCAAAGGACAACAAGCAATGGCATCCTTGCAGCTGCTGTCTCAGCCGATTGGAGCAGACA   840
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   261 ProProLysGluThrSerIleGlyIleLeuAlaAlaAlaValSerArgLeuGluGlnThr   280

Qy   841 CCAATGCCTATACATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAAT   900
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   281 ProMetProIleIlePheGlySerGlyThrValValThrValLeuGlnGlnLeuAlaAsn   300

RESULT 2
Q8R117
Q8R117
ID   Q8R117   PRELIMINARY;   PRT;   340 AA.
AC   Q8R117;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 37.8 kDa protein (Fragment).
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=LIVER;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC025830; AAH25830.1; -.
KW   Hypothetical protein.
FT   NON_TER   1   1
SQ   SEQUENCE   340 AA; 37800 MW; 585321EE72EFAD2E CRC64;

Alignment Scores:
Pred. No.:
Score: 1.09e-97
                       1261.50
Percent Similarity: 85.29%
Percent Identity: 69.12%
Best Local Similarity: 45.67%
Query Match: 11

Length: 340
Matches: 235
Conservative: 55
Mismatches: 49
Indels: 1
Gaps: 1

DB:

US-10-014-896-1 (1-1509) x Q8R117 (1-340)

Qy   901 GAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGACCACTTATA   960
          |||||||||||||||||||||||||||||||||||||||||||||||||------------
Db   300 -- 300

Qy   961 AGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGAGGACCACCAGGCACTC 1020
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 Glu--ValTyrGlyGluLysSerLeuAsnGlnCysAsnAsnGlnAspHisHisGlyThrH 320

Qy   1021 ACCATATTTCAAAGCAGGAGGGGTCAAGTTCAAATGTCATCCCCCCAGTGTGGCCCACAGTC 1080
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   321 IsHisIleGlnSerArgGlyValGlnCysHisProProSerGlyProGlyHisSerG   340

Qy   1081 AACTTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGAGGAACATT 1140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   340 InLeuProAspSerProTrpThrAspSerProArgGlyProArgThrHisGluGluHisC 360

Qy   1141 GTGGC 1145
          |||||
Db   360 ySGly 361

```

```
Qy   787 AAGGAGACAAGCATTGGCATCCTTGCAGCTGCTGTCAGCCGATTGGAGCAGACACCAATG 846
               ||||||||||||||||||||||||:::|||||||||||||||||||||||||||||||||
Db   101 LysGluThrSerIleGlyIleLeuSerAlaAlaValSerArgLeuGluGlnThrProMet 120

Qy   847 CCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAATGAGTTT 906
               ||| :::|||||| ||| :::   :::::: |||||||||||||||
Db   121 ProAsnMetPheGlyGlyGlyProLeuLysLysThrMetLysLeuLeuAlaAsnGluPhe 140

Qy   907 CCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGG 966
                  ||||||:::||||||:::||| ||| ||||||||| |||::::::||||||
Db   141 SerPheProIleAsnIleValLeuArgAsnLeuTrpLeuPheHisProIleValSerArg 160

Qy   967 TTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTCACCATA 1026
                  |||||||||||||||:::|||||||||::::::|||||||||||||||||||||:::
Db   161 IleMetGluArgAsnProIleThrAsnAlaLeuValArgThrThrThrAlaLeuThrMet 180

Qy   1027 TTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGCCACAGTCAACTTC 1086
               ||| ||||||:::||| |||||||||||||||:::||||||||||||:::|||
Db   181 PheAsnAlaGlyIleLysValAsnValIleProProLeuAlaGlnAlaThrIleAsnCys 200

Qy   1087 CGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGAAGAACATTGTGGCT 1146
               |||||||||||| ||||||||| ||||||||||||||| |||||| ||||||
Db   201 ArgIleHisProSerGlnThrValHisGluValLeuGluLeuValLysAsnThrValAla 220

Qy   1147 GATAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTTGACCCCCTCCCCGTCAGCCCTTCT 1206
               |||:::||||||||| ||||||||| :::|||:::|||||||||:::|||||||||
Db   221 AspAspArgValGlnLeuHisValLeuArgSerPheGluProLeuProIleSerProSer 240

Qy   1207 GATGACAAGGCCTTGGGCTACCAGCTGCTCCGCCAGACCGTACAGTCCGTCTTCCCGGAA 1266
               ||||||:::|||:::|||||||||||||||::::::|||::::::|||||||||||||||
Db   241 AspAspGlnAlaMetGlyTyrGlnLeuLeuGlnGluThrIleArgSerValPheProGlu 260

Qy   1267 GTCAATATTACTGCCCCAGTTACTTCTATTGGCAACACAGACAGCCGATTCTTTACAAAC 1326
               |||:::||| ||| ||| |||||||||:::||| ::: |||
Db   261 ValAspIleValValProGlyIleCysIleAlaAsnThrAspThrArgHisTyrAlaAsn 280

Qy   1327 CTCACCACTGGCATCTACAGGTTCTACCCCATCTACATACAGCCTGAAGACTTCAAACGC 1386
               :::||| |||:::||||||||| |||:::   ::: |||:::||||||
Db   281 IleThrAsnGlyMetTyrArgPheAsnProLeuProLeuAsnProGlnAspPheSerGly 300

Qy   1387 ATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAAATTCATC 1446
               :::||||||:::|||||||||:::||||||||| |||::: |||||||||||||||
Db   301 ValHisGlyIleAsnGluLysValSerValGlnAsnTyrGlnAsnGlnValLysPheIle 320

Qy   1447 TTTGAGTTGATTCAGAATGCTGACACAGACCAGGAGCCAGTTTCTCACCTGCACAAACTG 1506
               |||||| |||||||||||||||||| :::||||||||| |||||||||:::|||
Db   321 PheGluPheIleGlnAsnAlaAspThrTyrLysGluProValProHisLeuHisGluLeu 340

RESULT 3
Q8XSI7
ID   Q8XSI7   PRELIMINARY;   PRT;   510 AA.
AC   Q8XSI7;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical transmembrane protein RSp0487.
GN   RSP0487 OR RS00365.
OS   Ralstonia solanacearum (Pseudomonas solanacearum).
OG   Plasmid megaplasmid.
OC   Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC   Ralstonia.
OX   NCBI_TaxID=305;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=GMI1000;
RX   MEDLINE=21681879; PubMed=11823852;
RA   Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA   Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA   Chandler M., Choisne N., Claudel-Renard C., Cunnac S., Demange N.,
RA   Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA   Siguier P., Thebault P., Whalen M., Wincker P., Levy M.,
RA   Weissenbach J., Boucher C.A.;
RT   "Genome sequence of the plant pathogen Ralstonia solanacearum.";
```

```
RL   Nature 415:497-502(2002).
DR   EMBL; AL646079; CAD17638.1; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
KW   Plasmid; Hypothetical protein; Complete proteome.
SQ   SEQUENCE   510 AA;  54350 MW;  F3FEDB6995BC9C7C CRC64;

Alignment Scores:
Pred. No.:              3.96e-57      Length:        510
Score:                  781.00        Matches:       177
Percent Similarity:     60.14%        Conservative:  90
Best Local Similarity:  39.86%        Mismatches:    159
Query Match:            28.28%        Indels:        18
DB:                     16            Gaps:          11

US-10-014-896-1 (1-1509) x Q8XSI7 (1-510)

Qy         172 GCCATCCAGATTCCAACAGTG---ACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTG 228
               ||||||:::      |||:::   :::|||:::      :::   |||
Db          68 AlaIleArgAlaArgThrIleSerSerPheThrAspAlaGlnGlnAsnAlaAspGlnPhe 87

Qy         229 GCTGAGTTCGGAAAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATC 288
                  :::            :::         :::|||         :::   :::   :::
Db          88 ArgGlnLeuHisAlaMetLeuAlaAlaArgTyrProLys------AlaHisAlaAlaMet 105

Qy         289 CAGCATGAAGTCGTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTCGGACCCCAGC 348
               |||   |||   |||   ::::::|||   |||:::|||   :::|||||||||||||||
Db         106 GlnArgGluProValGlyAspPheSerLeuLeuTyrThrTrpLysGlySerAspProSer 125

Qy         349 TTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGTGCCT---GCCCCT------GAA 399
               |||:::|||   |||||||||||||||   ||||||||||||   ||||||      |||
Db         126 LeuLysProIleLeuLeuMetAlaHisGlnAspValValProValAlaProGlyThrGlu 145

Qy         400 GAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGG 459
                     |||      |||||||||:::|||:::   :::||||||:::::::::||||||
Db         146 GlyAspTrpThrGluProProPheAlaGlyValValLysAspGlyMetValTrpGlyArg 165

Qy         460 GGCACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTGCTG 519
               |||      |||||||||   :::::::::|||   ::::::|||   |||||||||
Db         166 GlyAlaTrpAspAspLysGlyAsnLeuIleAlaGlnMetGluAlaAlaGluLeuLeuAla 185

Qy         520 ATCAGGAAGTACATCCCCCGAAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTCA 579
                        :::   |||||||||:::         :::   |||   |||||||||
Db         186 AlaSerGlyPheArgProArgArgThrIleHisPheAlaPheGlyAlaAspGluGluVal 205

Qy         580 TCAGGG---ACAGGGGCTCAGAGGATCTCAGCCCTGCTACAGTCAAGGGGCGTCCAGCTA 636
                  |||      ||||||   ||||||:::|||||||||:::|||||||||   ||||||
Db         206 GlyGlyGluArgGlyAlaAlaArgIleAlaAlaLeuLeuLysSerArgGlyGluGlnLeu 225

Qy         637 GCCTTCATTGTGGACGAGGGGGGCTTCATCTTGGATGATTTCATTCCTAACTTCAAGAAG 696
               ||||||::::::|||||||||      |||   :::      :::|||         |||
Db         226 AlaPheValIleAspGluGlyLeuLeuIleThrGluGlyValValProGlyLeuThrLys 245

Qy         697 CCCATCGCCTTGATTGCAGTCTCAGAGAAGGGTTCCATGAACCTCATGCTGCAAGTAAAC 756
               |||   |||||||||   |||:::|||||||||   :::::::::   |||:::::::::
Db         246 ProAlaAlaLeuIleGlyValThrGluLysGlyPheLeuSerValAlaLeuLysLeuSer 265

Qy         757 ATGACTTCAGGCCACTCTTCAGCTCCTCCAAAG------GAGACAAGCATTGGCATCCTT 810
                  |||   ||||||||||||   ||||||         |||::::::|||   ::::::
Db         266 AlaThrProGlyHisSerSerMetProProAlaProGlyGluSerAlaIleAlaMetMet 285

Qy         811 GCAGCTGCTGTCAGCCGATTGGAGCAGACACCAATGCCTATCATATTTGGAAGCGGGACA 870
               :::||||||:::      |||:::         :::|||               |||
Db         286 SerAlaAlaLeuLysHisLeuAspAspGlnGlnLeuProAlaGlyIle---ArgGlyVal 304

Qy         871 GTGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCCTTCCCTGTCAATATAATCCTG 930
                        :::   :::   ||||||   |||                  :::   |||
Db         305 ThrArgGluMetPheGluThrLeuAlaProGluMetArgGlyLeuAsnArgValAlaLeu 324

Qy         931 AGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACC 990
               ||||||   ||||||||||||||||||:::   :::   ::::::   :::|||   |||
```

RESULT 4
Q8RTT1

ID   Q8RTT1   PRELIMINARY;   PRT;   488 AA.
AC   Q8RTT1;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   M20/M25/M40 family peptidase, putative.
GN   EBAC000-65SD09.52.
OS   uncultured proteobacterium.
OC   Bacteria; Proteobacteria; environmental samples.
OX   NCBI_TaxID=153809;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21822632; PubMed=11832943;
RA   Beja O., Suzuki M.T., Heidelberg J.F., Nelson W.C., Preston C.M.,
RA   Hamada T., Eisen J.A., Fraser C.M., DeLong E.F.;
RT   "Unsuspected diversity among marine aerobic anoxygenic phototrophs.";
RL   Nature 415:630-633(2002).
DR   EMBL; AE008919; AAL76395.1; -.
SQ   SEQUENCE 488 AA; 53345 MW; AD3F7179D9B83486 CRC64;

Alignment Scores:
Pred. No.:
Score: 639.50
Pred. No.:   3.3e-45
Score: 639.50
Percent Similarity: 52.36%
Percent Similarity: 35.06%
Best Local Similarity: 23.15%
Query Match: 2

                        Length: 488
                        Matches: 156
                        Conservative: 77
                        Mismatches: 195
                        Indels: 17
                        Gaps: 10

US-10-014-896-1 (1-1509) x Q8RTT1 (1-488)
```

```
Qy   157 GAGGCGGCTGAAAAGCTGCCATCCACGATTCCAAACACAGTTGACTTTT---AGCTCTGAGAAGTCC 213
         ||| :: :: ||||| ||| ::||| |||:::|||||||| ::: |||::::::|||:::
Db   49 AsnArgLeuAlaThrAlaValArgPheLysThrValSerSerGlnAspThrSerLysIle 68
```
(...detailed nucleotide/protein alignment rows follow...)

```
Qy  1273 ATTACTGCCCAGTTACTTCTATTGGCAACACAGCAGCCGGATTCTTACAAACCTCACC 1332
              |||   |||||:::|||   ||| |||||||
Db   420 LeuAlaIleProSerLeuLeuGlnAlaThrThrAspThrArgHisTyrValAsnLeuAla 439

Qy  1333 ACTGGCATCTACAGGTTCTACCCCATCACTACACAGCCTGAAGACTTCAAACGGCATCCAT 1392
                 |||||||:|||            |||  |||     |||   |||||||
Db   440 LysAspGlnTyrArgPheHisGlyAsnSerIleAspAlaSerGlnAlaAlaArgSerValHis 459

Qy  1393 GGACTCAACGAGAAAATCTCAGTCCAAGCTATGAGACCCATGAAATTCATCTTTGAG 1452
         ||||||                  :::||:||||
Db   460 GlyThrAsnGluTyrIleSerGluArgSerTyrAsnAsnAlaIleAlaValAlaArgGly 479

Qy  1453 TTGATTCAGAATGCT 1467
         |||       |||
Db   480 MetLeuLysGlyAla 484
```

RESULT 5
Q9AA70

```
ID   Q9AA70       PRELIMINARY;      PRT;   471 AA.
AC   Q9AA70;
DT   01-JUN-2001 (TREMBLrel. 17, Created)
DT   01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE   M20/M25/M40 family peptidase.
GN   CC0736.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RL   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005748; AAK22721.1; -.
DR   TIGR; CC0736; -.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
KW   Complete proteome.
SQ   SEQUENCE   471 AA;  47867159BBE0CDEF CRC64;
```

Alignment Scores:

```
Pred. No.:
Score:                   591.00
Percent Similarity:      52.218
Best Local Similarity:   33.418
Query Match:             21.408

                                        3.99e-41
                                        16

Length:          471
Matches:         151
Conservative:    85
Mismatches:      184
Indels:          32
Gaps:            12
```

US-10-014-896-1 (1-1509) x Q9AA70 (1-471)

```
Db   90 TrpThrGlySerAsnProAlaLeuHisProIleValLeuMetAlaHisGlnAspValVal 109
Qy  157 GAGGCCGCTGAAAGGTGCCATCCAGAATTCCAACAGAGCACTTTTAGCTCTGAGAAGTCCAAT 216
             |||            |||||:::             |||   ||:||::::
Db   388 ------------------CCTGCCCTGAAGAGAGGCTGGGAGGCGCCCCATTCTCTGGGTTGGACGT 438

Qy   34 GAActtcacaGAGGagcctGAAGATCCAActtcaccATCACCATCGCCCATCTCACCATCAAGAGCCTTGGGCTACCAGCTGCCAACAC 53
Db   110 ProValThrProGlySerGluGlyGlnTrpThrHisProProPheAlaGlyValValAla 129

Qy  217 ACTACAGCCCTGGCTCAGTTCGGGAAAA---------TACAtTCATAAAGtCTTCCTACA 267
                 ||||:::||:|||||            ||||||||    ||:|||
Db  130 AspGlyLysValTrpGlyArgGlyAlaAlaAspAlaGlySerLeuValThrIlePhe 149

Qy   54 GlnPro----------AlaGluTrpAspLysLeuHisAlaTrpLeuGlnThrThrTyrProGln 71
Db  268 ------------------GGAACGGTGCAGAGTGGTCAACGAACATTGTGGCTGTGATAAACAGA 1155
```

(Remaining alignment blocks continue pairing query nucleotide sequence with database protein/nucleotide sequence.)

```
Qy    1432 CAAGTGAAATTCATCTTTGAGTTGGATTCAGAATGCT 1467
            ||||:||||| || ||||| ||||||||||:||| |||
Db     457 MetValArgPheTyrGlnArgLeuValGluThrAla 468

RESULT 6

ID    O13968
AC    O13968;
DT    01-JUN-1998 (TrEMBLrel. 06, Created)
DT    01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Putative carboxypeptidase s precursor (EC 3.4.17.4) (YSCS) (GLY-X
DE    carboxypeptidase).
GN    SPAC24C9.08.
OS    Schizosaccharomyces pombe (Fission yeast).
OC    Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC    Schizosaccharomycetales; Schizosaccharomycetaceae;
OC    Schizosaccharomyces.
OX    NCBI_TaxID=4896;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Murphy L., Harris D., Barrell B.G., Rajandream M.A., Wood V.;
RL    Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
CC    -!- FUNCTION: NECESSARY FOR USE OF CERTAINS PEPTIDES AS SOLE NITROGEN
CC        SOURCE. MAY ALSO CLEAVE INTRACELLULARLY GENERATED PEPTIDES TO
CC        RECYCLE AMINO ACIDS FOR PROTEIN SYNTHESIS (BY SIMILARITY).
CC    -!- CATALYTIC ACTIVITY: PEPTIDYLGLYCINE + H(2)O = PEPTIDE + GLYCINE.
CC    -!- PATHWAY: NITROGEN METABOLISM.
CC    -!- SUBCELLULAR LOCATION: LYSOSOME-LIKE VACUOLES (BY SIMILARITY).
CC    -!- PTM: GLYCOSYLATED (BY SIMILARITY).
CC    -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M20A; ALSO KNOWN AS THE
CC        ARGE/DAPE/ACY1/CPG2/YSCS FAMILY.
DR    EMBL; Z98601; CAB11265.1; -.
DR    MEROPS; M20.002; -.
DR    InterPro; IPR002933; Peptidase_M20.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfam; PF01546; Peptidase_M20; 1.
DR    PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
KW    Hypothetical protein; Hydrolase; Carboxypeptidase; Glycoprotein;
KW    Signal.
FT    SIGNAL        1      ?       POTENTIAL.
FT    CHAIN         ?      596     PUTATIVE CARBOXYPEPTIDASE S.
FT    CARBOHYD    118      118     N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD    466      466     N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD    541      541     N-LINKED (GLCNAC...) (POTENTIAL).
FT    CARBOHYD    555      555     N-LINKED (GLCNAC...) (POTENTIAL).
SQ    SEQUENCE    596 AA;   66606 MW;   D97A39307443364C1 CRC64;

US-10-014-896-1 (1-1509) x O13968 (1-596)

Alignment Scores:
Pred. No.:
Score:                4.47e-36
Percent Similarity:   531.50
Best Local Similarity: 40.75%
Query Match:          25.30%
DB:                   19.24%
                      3

                      Length:         596
                      Matches:        149
                      Conservative:   91
                      Mismatches:     212
                      Indels:         137
                      Gaps:           13
```

```
Qy     28 GCCCTGGTGGCTATGCTGCTCCCTAGTTTTCCCTACCGTCTCCCAGATCGANGGGCCCGAGG 87
Db     40 AlaPheValGlyPheLeuIleLeuIlePheValTyrTyrLeuLeuArgGlyGlySerAsn 59

Qy     88 AGCCGGGGACCATCAAAAGGCCGTCCGGAACCCCTTCTCAGTTCAGCAAAGAGAAGGAACGGCGTC 147
Db     60 AspAsnAspGlyGlnGluMetSerHisSerProGlySerCysMetAspSerGluSerAla 79

Qy    148 GCGATGAAAAGAGGGCCCTGAAA 168
Db     80 AlaValSerThrSerAlaLysCysTyrIleProProValLeuThrProAlaLysGluPro 99

Qy    168 ---------------------------------------------------------------- 168
```

```
Db    100 LysLeuGlyAspAspValSerGlyIleAspTyrIleArgSerProGluPheAsnAsp 119

Qy    169 -------------TCTGAGAAGTCCAATACTACAGCCGTCGGGCTGGTGGAGTTCGGAAAATACATTCATAAA 255
            ::: ||||:::||::::::||||||
Db    120 SerLeuValArgPheGlyLnGlyLeuLeuArgIleLeuProThrValCysTyrGlyMetGly 139

Qy    202 AspValGlyAspAspAspArgPheAspIlePheAlaValPheGlnLnAspLysValArgGlu 159
Db    140 AspValGlyAspAspAspArgPheAspIlePheAlaValPheGlnLnAspLysValArgGlu 159

Qy    256 GTCTTTCCTACAGTGGTCAGCGACCACCAGCTTTATCCAGCCATGAAGTGTGGAAGAGTATAGC 315
Db    160 LeuTyrProAsnIlePheLysLys------LeuLysValGluTyrValAsnThrTyrGly 177

Qy    316 CACCCTGTTCACTATCCAAGGCTTCGGACCCCCAGCGTTGCACGCCTACCTGCTTGATGGCTCAC 375
Db    178 LeuLeuIleThrLeuGluGlySerAsnLysAspLeuLysProLeuValLeuMetGlyHis 197

Qy    376 TTTGAGTGGTGGCCT 426
Db    198 GlnAspValProValAsnGlnAlaSerLeuAspArgTrpIleTyrPheProProSer 217

Qy    427 GGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGCACACTGGACGACAAGAACTCTGTG 486
Db    218 AlaThrTyrHisAsnGlyHisValTyrSerArgGlyAlaAlaAspAspLysAsnSerVal 237

Qy    487 ATGGGATTACTGCAGGCGTCCCTTGGAGGTCCCTGCTGATCAGAGAAGTACATCCCCGAAGATCT 546
Db    238 ValAlaIleLeuGluValLeuGluAlaLeuGluAlaIleSerAspTyrLysProGlnThr 257

Qy    547 TTCTTCATTTCTCTGGGCCATGATGAGGAGTCATCAGGG---ACAGGGGCTCAGAGGATC 603
Db    258 ValIleAlaSerPheGlyPheGluAspGluValSerGlyTyrArgGlyAlaLeuProLeu 277

Qy    604 TCAGCCCTGCTACAGTCAAGG----------GGCGTCCAGCTAGCCTTCATGTGGGAC 651
Db    278 AlaHisLysLeuTyrGluArgTyrGlyLysAspGlyVal-----AlaLeuIleLeuAsp 295

Qy    652 GAGGGGGGCTTCATCTTGGATGAATTCATTCCTAACTTCAAGAAGGACCCATGCCTTGATT 711
Db    296 GluGlyGlyPheThrIleAspGluPheIleProAsnPhe----------AlaThrVal 311

Qy    712 GCAGTCTCAGAGAAGGGGTTCCATGAAACCTCATGCTGCAAGTAAACATGACTTCAGGGCAC 771
Db    312 CysValAlaGluTyrGlyTyrMetAspAspValHisLeuLysLeuThrProGlyHis 331

Qy    772 TCTTCAGCTCCTCCAAAAGGAGACAGAGAAGCANTTGGCATCCTTGCAGCTGCTGGTCGGCGATTG 831
Db    332 AlaSerIleProProProHisThrAsnIleGlyLeuMetSerLysLeuValThrGlnIle 351

Qy    832 GAGCAGACACCAAATGCCTATCANTATTTGGAAGCGGGACAGATGGTGGTGACTGTATTGCAGCAA 891
Db    352 GluGlu------------------------------------------------------ 353

Qy    892 CTGGCAAATGAGTTTCCCTTCCCGTCAAATATAATCCTGAGCAACCCATGGGCTATTTGAA 951
Db    354 ---------ProPheGlyGluThrPheGluAsnPro------------PheTyr 366

Qy    952 CCACTTATAAAGCAGGTTTATNGGAG 975
Db    367 ThrThrLeuGlnCysPheAlaGluAsnSerAlaAspMetAspAspAsnLeuArgGlnLeu 386

Qy    976 -----------------AGAAATCCCTTAACCAAT 993
Db    387 IleLysSerGlyAspThrGlyLysMetThrAspLeuPheSerLysSerArgLeuTyrArg 406

Qy    994 GCAATAAATCAGGACCACCACGGCACTCACCATATTCAAAAGCAGGGGTCAAGTTCAACTC 1053
Db    407 TyrPhePheGluThrSerIleAlaValAspValIleAsnGlyGlyValLysValAsnAla 426

Qy   1054 ATCCCCCAGTGGGGCCCAGCCCACAACTTCCGGATTCAACTCTCGGACAGCAGTCCAA 1113
            ||||| ||:::                  :::                  ::::::
```

```
RESULT 7

Q8RNM5

ID    Q8RNM5           PRELIMINARY;     PRT;    469 AA.
AC    Q8RNM5;
DT    01-JUN-2002 (TrEMBLrel. 21, Created)
DT    01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Zn metalloprotein.
OS    Legionella pneumophila.
OC    Bacteria; Proteobacteria; gamma subdivision; Legionellales;
OC    Legionellaceae; Legionella.
OX    NCBI_TaxID=446;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Rankin S., Li Z., Isberg R.R.;
RT    "Macrophage Induced Genes of Legionella pneumophila: Protection from
RT    Reactive Intermediates and Solute Imbalance During Intracellular
RT    Growth.";
RL    Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; AF480917; AAM00645.1; -.
SQ    SEQUENCE  469 AA;  51844 MW;  E8F1E46F0D99CF1B CRC64;

Alignment Scores:
Pred. No.:                       6.82e-28
Score:                           434.00
Percent Similarity:              45.60%
Best Local Similarity:           28.00%
Query Match:                     15.71%
DB:                              2

US-10-014-896-1 (1-1509) x Q8RNM5 (1-469)

Length:          469
Matches:         140
Conservative:    88
Mismatches:      208
Indels:          64
Gaps:            17
```

```
Qy    1177 GCCTTTGACCCCGTCAGCCCTTCTCGATGACAAGGCCGTTGGGCTACCAGCTGCTC 1236
           ||||| :::    ||||||  :||       |:||    ||||||:   ||||||
Db     372 LysLeuSerArgGlyProSerSerProLeuSerProGluIleLeu-----LysThrIle 389

Qy    1237 CGCCAGACCCTACACTGCCTCCGTGTCCCGAAGTCAATATTACTCGCCCAGTTACTCGTATT 1296
           ||| ||||||||||  ::|||||||||||||
Db     390 ThrGlnLeuthrGlnArgTyrTrpProGluValPro---ThrIeProIleMetValThr 408

Qy    1297 GGCAACACAGACAGCCGATTCTTTACAAACCTCACCACTGGCACTACCAGGTTCTACCCC 1356
           |||:    ::::        :::
Db     409 GlyAlaThrAspGlyArgPheLeuArgSerValGlyIleProThrTyrGlnValMetGly 428

Qy    1357 ATCCACATACCAGGCTGAAGACTTCAAACGCATCCATGAGCTCAACGAGAAAATCTCAGTC 1416
           :::|||||||||  |||| ||||||::||||||  ||||||||:::||||||||||
Db     429 LeuPheLeuAspArgAspAspAspPhe---ArgAlaHisGlyIleArgHisGluArgIle 447

Qy    1417 CAAGGCCTATGAGAGACCCAAGTGAAATTCATCTTTGAGTTGATTCAGAAATGCTGACACAGAC 1476
           ||:|:|||||:||||||||||:::|||  ||||:::|
Db     448 GluSerPheTyrGluAlaHisAlaAlaPheLeuTyrAspLeuValLysGlnLeuSerSerAsn 467
```

RESULT 8

```
US-10-014-896-1 (1-1509) x Q92BI7 (1-443)

Alignment Scores:
Pred. No.:
Score:                8.63e-22
Score:                361.50
Percent Similarity:   45.638
Best Local Similarity: 29.268
Query Match:          13.098
DB:                   16

Length:           443
Matches:          134
Conservative:     75
Mismatches:       181
Indels:           68
Gaps:             20
```

```
Qy     156 AGAGCGGCTGAAAAGGTGCCAACCCAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAA 215
           |||||        :::     ||| |||    :::           :::||||||  :::
Db      34 ArgGlyGlyGlyAspCysGln-GluArgProAlaAlaGluTyrAlaAlaArg----- 51

Qy     216 TACTACAGCCCTGCCTGGCTGAGTTCGGAAAATACATTCATAAAGTCTTCCTACAGTGGTCAG 275
           ||||   ||||:||||||||||||||::          :::    |||||:::
Db      52 ----------LeuLAlaGluAlaGly-------------IleGluProThrLeuLeuGl 63

Qy     276 CACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCCACACCTGTTCACTATCCAAGG 335
           ||||||        :::                 ||||:: ::||  |||::||
Db      63 uArgThrAlaGlyArgThrAsnValValAlaArg-------------MetAspAlaMetLysAsn 77

Qy     336 CTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGTGCCTGCCC 395
           |||| :|| :||||||||  ||:||||:||||||||||:    :::|||||
Db      77 yThrAspProSerAlaAspValAsnProAlaLeuLeuValHisGlyHisLeuAspValValProAlaLeuLeu 97

Qy     396 TGAAGAAGGCTGGGAGGGTGCCCCCATTCTTCTGGGTTGGAGAGTCATCTATGG 455
           ||| |||||||||||||::|||||     :::|||           ||| |||
Db      97 uAlaLaAlaAspTrpSerValHisProPheSerGlyGluIleLeuAlaArgHisAl 117

Qy     456 TCGGGGCACACTGGACGACGCAAAGAACTCTGTGATGGCATTACTGCAGGCCTTCCT 515
           ||| ||||   ||| ||| |||    :::|||             :::|||||
Db     117 yArgGlyYAlaAlaValAspAsprMetLysAsn---MetAspAlaMetLysAsn 132

Qy     516 GCTGATCAGGAGGAGTACATC------------CCCCGAAGACTTTCTTCATTTCT 560
           |:::||       |||:::                |||||||||||||||:||
Db     132 aValValArgAspTrpPAlaAAsnGlyValArgGlyValArgGlyArgAspValValIleAlaPh 152

Qy     561 GGGCCATGAATGAGGAGTCATCA------------GGGACAGGG----GCTCAGGAGGACTC 605
           |||||||:||||||||||:::|||             |||||||      ||:|||||||||
Db     152 eThrAlaAspGluGluAlaSerAlaGluAspGlySerGlyPheLeuAlaAspArgHisAl 172

Qy     606 AGCCCTGCTACAG-----------TCAAGGGGGGTCCAGCTAGCCTTCATTGTGGGACGAGGGGG 659
           |||||||||||             :::|||||||      :::||||||||||:
Db     172 aAlaLeuPheGluGlyCysThrGluGlyVAlSer------GluSerGlyyAl 187

Qy     660 CTTCCATCTTGGATGAT--------TTCATTCCTAACTTCAAGAGAAGCCCATCGC 704
           ||||||||||:::|||            :::||||||||||||||:::||||||
Db     187 aPheThrPheHisAspThrAAlaGlyyAlaGlyVArgGlnPheThrTyrPro------ 200

Qy     705 CTTGATTGGCAGTCTCAGAGAAAGGGGTTCCATGAACCTCATGCGCAAGTAAACATGACTTC 764
           |||||||||||         :::||||||||:::|||         :::|||||||||||
Db     201 ----IleAlaAlaGlyyArgGlyYTrpLeuUyrSerLeuThrArgLeuArgGlyyArgAl 219

Qy     765 AGGCCACTCTTCAGCTCCTCCAAAGGAGAGCAGCACCATTGGGAATCCTTGCAGCTGCTGTCAG 824
           |||||||||||||||      :::||||||||||||||||||||||||||||||||
Db     219 aGlyyHisGlySerLysValAsnArgGluAsnArgLeuLeuAlaGluAlaLeuTrp 239

Qy     825 CCGATTGGAGCAGCAGACACCAAATGCCTATCAATATTGGNAACGGGACACAGTGGTGGACTGTATT 884
           :||:::|||  ||:::|||||||||||||||||    :::|||           |||
Db     239 rArgyIleGlyyAspHisAlaTrpProLeuuArgyLeu---ThrProThrValArgAlaAlaLe 258

Qy     885 GCAGCAACTGGCCAAATGAGTTGTCCCTGTCCCGTCCAATATATNTCGAGCAACCCATGGCT 944
           ||||       :::|||||||||||||||||||||||||    :::|||||   :::|||
Db     258 uThrGluuIleAlaGlyyValTyrGlyyIleGluuThrAsp-----LeuuSerAsp- 273

Qy     945 ATTTTGACCACTTATAAGCAGGTTTAYGGAGAGAAANCCCTTAACCAATGCAATAATCAG 1004
           |||:::|||||||||||  :::||||           :::|||||||||||||||
Db     274 -ValAspAlaLeuuLeuAspLysLeuuAspGlyyGlnAlaGlyyLysLeuuValGlyy 293

Qy    1005 GACCACACGGCACTCACCATATTCAAAGCAGGGCGTCAAGTTCAATCATCCCCCAGT 1064
           :::||| ||||||||||||                  |||||   ||||||||||
Db
```

```
RESULT 9
Q9A3G5
ID   Q9A3G5          PRELIMINARY;      PRT;   474 AA.
AC   Q9A3G5;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Peptidase, M20/M25/M40 family.
GN   CC3239.
OS   Caulobacter crescentus.
OC   Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC   Caulobacter.
OX   NCBI_TaxID=155892;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 19089 / CB15;
RX   MEDLINE=21173698; PubMed=11259647;
RA   Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA   Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA   Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA   DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA   Kolonay J.F., Smith J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA   Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA   Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT   "Complete genome sequence of Caulobacter crescentus.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR   EMBL; AE005988; AAK25201.1; -.
DR   InterPro; IPR002106; AatRNA_ligaseII.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   474 AA;  50764 MW;  D9E74458A753D689 CRC64;
```

Alignment Scores:

Pred. No.:	1
Score:	4.28e-20
Score:	341.50
Percent Similarity:	43.98%
Percent Identity:	27.35%
Best Local Similarity:	12.36%
Query Match:	16

Length:	474
Matches:	125
Conservative:	76
Mismatches:	235
Indels:	21
Gaps:	12

DB:

```
us-10-014-896-1 (1-1509) x Q9A3G5 (1-474)

Qy   118  CCTTCTCAGTTCAGCAAAGAGAGGAACACCGTTCCGGATGAGAAAGGGCCCATC  177
          ::: :::::  :::::   ::: :    ::  :   : :::    ::   :::
Db    26  ProProLysThrValArgProAspGlnLeuAlaMetArgAspLeuThrTyrLysGluGluVal  45

Qy   178  CAGATTCCAACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCTGCGACCTGGGTAGTTC  237
          ::::: ::   ::         ::   :::   :::::    ::::::   :::
Db    46  GluIleAsnThrThrLeu-----SerValGlySerCysThrAlaAlaSerGluAlaMet  63

Qy   238  GGAAAATACATTCATTAAAGTC---TTTCCTACAGTGCATGGTGCCAGTGGTCACCACCAGCAT  294
          :::::::  :::  ::::::   : :::::::   :: ::   ::: ::::  :::
Db    64  GlyAlaArgLeuLeuAlaAlaGlyPheProGluAlaAspValArgValValGluPro  83

Qy   295  GAAGTCGTGGAAGAGTATAGCCACCGTGTTCACTACTACTACACACCAGACCCTTGCAG  354
          :::: :::  ::::::::   ::::::   :  ::::::   ::::::: :::
Db    84  LysHisProArgGluGlyAsnLeuValAlaValLeuArgGlyThrArgThrLys  103

Qy   355  CCCTACCTGCTGCATGGCCTCACTTTGATGTGGTGCCCCTGAAGAGGCTGGGAGGTG  414
          ::::::::  ::::  :::   ::::::: :::::   ::::::::: ::::::
Db   104  ProMetLeuLeuLeuHisIleAspValValGluAlaLysSerGlyGluTrpProArg  123

Qy   415  CCCCCATTCTCTGGGTTGGAGGCGTTGGATGGCGTGCATCATCTATGGTCGGGGCCACACTGGCGACGAC  474
          ::::::::: ::: ::::::::  :::::::: ::   :  :::   ::::::::::: :::
Db   124  AspProPheLysLeuValGlnGluAsnGlyTyrProPheTyrGlyArgGlyThrSerAspAsp  143

Qy   475  AAGAACTGTGTGATGGCATTACTGCGAGCCGCTCCCTGCCTGATCAGGAGAAGCCATCGCC  534
          :::::::: ::  ::::::::::: :::::::  :::  ::::::::::::
Db   144  LysAlaGlnAlaAlaIleLeuTrpValAspSerLeuIleArgLeuLysGlnAlaGlyPheLys  163

Qy   535  CCCGAAGATCTTTCTTCGATTTCTCTGGGCCATGATGAGGGAGGTTCATCGACAGCAGGGGCT  594
          ::: :::::::   ::::::::: ::::::::::::::::: ::   ::::::::
Db   164  ProLysArgAspSerIleLeuLysMetAlaLeuLeuThrCysGlyCysGluSerGluGlyThrTyrArgAsnGly  183

Qy   595  CAGAGGATCTCAGCCCTGCCTACAGCCAGTCAAGGGGC-----GTCCAGCTAGCCTTCATTGTG  648
          ::::::::: :::::: :::   :::::::::::::::     :: ::::::::::: ::::::
Db   184  -----IleGluSerLeuLeuLysAsnHisArgProLeuValAlaValAspAspAlaLeu  201

Qy   649  GACGGAGGGGGGCTTC---ATCTTGGATGAGATTCATTCATCCTAACTTCAAGAAGCCCATCGCC  705
          :::::::::::::::   ::::::::::::::: ::::::
Db   202  AsnGluGlyAlaSerGlyLeuLeuLeuAspGlu------GlnGlyLysAlaValMetLeu  218

Qy   706  TTGATTGCAGTTCTCAGACAGAAGGGTTCCATGAACCTCAAGTAAACAATGACTTCA  765
          :::  ::: :::::::::::::::::: :::::: ::::::::::    :::
Db   219  GluValGlnAlaGlyLeuGluLysValAlaThrTyrGlnAspPheThrLeuThrValThrThrAsnProGly  238

Qy   766  GGGCCACTCTTCAGCTCCTCCCAAAGGAGGACAAGCAATTGGCATCCTTGCAGCGCTGTCAGC  825
          :::  :::: ::::::: : ::::::   :: ::::::::::: :::::
Db   239  GlyHisSerSerArgProValSerProAlaAlaIleTyrGlnLeuSerAlaAlaLeuAsp  258

Qy   826  CGATTGGCAGCAGACAACAATGCCTATACATATATTGGAGCGGCGAACAGTGGTGGACTGRATG  885
          :::  :::::   :: :::::: :::   :   :::  :::::::::::::
Db   259  ArgIleGlyAlaTyrGlnPheProProArgPheAsnAspAlaThrArgGlyTyrPheThr  278

Qy   886  CAGGCAACTGGCAAATGAGTTTCCC------TTCCTGTCAATATAATCCTGAGGAAC  936
          :::::::::::: :::::::::::     ::::::  ::::::::
Db   279  GlnMetGlnAsnLeuValAlaArgLysThrHisLeuValAlaAsp  298

Qy   937  CCATGGCTATTTGAACCACTATATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCA  996
          :::::  :::::::::::::::::::::  ::::::  :::::::::
Db   299  ------ValAsnAspProAlaAlaAlaLeuIleLeuThrLysAspArgThrTrpAsnSer  316

Qy   997  ATAATCAGGACACCACCACGGTCCACTCCACCATATTCAAAAGCTTCAATGTCATC  1056
          :::::::::::::::::::: ::::: ::::::::::::::
Db   317  MetLeuArgThrThrCysValAlaThrMetValSerAlaGlyHisAlaProAsnAlaLeu  336

Qy  1057  CCCCAGTGTGGCCCAGGCCACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAG  1116
          :::     ::::::::::: ::::  :::::: :::::::::
Db   337  ProGlnArgAlaThrAlaAsnIleAsnCysArgIleLeuProGlyThrProIleAspGlu  356

Qy  1117  GTCCTAGAACTCACGAAGAAAACATTGTGGCTGATAAAACAGAGTCCAGTTCCAATGTGTTGAGT  1176
          :::     :::::::::::::::::::
Db   357  ValLysAlaLysLeuThrGluLeuAlaAlaAlaAspProAlaValAla------ValThrLeu  374
```

DB: 9

```
Qy   1177  GCCTTTGACCCCCTTCAGCCCTTCTGATGACAAGGCCGTTGGGCTACCAGCTG---  1233
           |||                    |||||                    |||||
Db    375  AlaHisSerSerLysProAlaSerProProProLeuThrProAlaIleMetAlaPro  394

Qy   1234  CTCCGCCAGACGCTACAGTCCGTTCTTCCCGGAAGTCAATATTACTGCCCCAGTTACTTCT  1293
Db    395  IleGlnLysAsnAlaAlaLysLeuTrpProGlyValProIle---LeuProValMetSer  413

Qy   1294  ATTGGCAACACAGACGCCGATTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTAC  1353
           |||||||||||:::                    |||
Db    414  ThrGlyAlaThrAspAlaValHisThrSerAlaLeuGlyIleProThrTyrGlyValThr  433

Qy   1354  CCCAATCTACATACAGCCCTGAAGACTTCAAACGGCATCCATGGACTCACGAGAAAATCTCA  1413
Db    434  GlyLeuPheHisGlyProGluGly---ThrGlyValAlaHisGlyLeuAsnGluArgMetArg  452

Qy   1414  GTCCAAGGCCTATGAGACCCAAGTGAAAATTCATCTTTGAGTTGATTCAGAAT  1464
           ||:||||:::::|                    |||
Db    453  ValLysSerLeuTyrGluGlyLysArgAspPheLeuHisGlyLeuGlnAsp  469

RESULT 10

ID    Q9CC46;                PRELIMINARY;    PRT;   467 AA.
AC    Q9CC46;
DT    01-JUN-2001 (TrEMBLrel. 17, Created)
DT    01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT    01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE    Possible peptidase.
GN    M1288.
OS    Mycobacterium leprae.
OC    Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC    Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX    NCBI_TaxID=1769;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=TN;
RX    MEDLINE=21128732; PubMed=11234002;
RA    Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA    Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA    Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA    Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA    Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA    Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA    Rutter S., Seeger K., Simmonds M., Skelton J., Squares R.,
RA    Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA    Barrell B.G.;
RT    "Massive gene decay in the leprosy bacillus.";
RL    Nature 409:1007-1011(2001).
DR    EMBL; AL583921; CAC31669.1; -.
DR    Leproma; M1288; -.
DR    InterPro; IPR001261; ARGE_DAPE_CPG2.
DR    InterPro; IPR002933; Peptidase_M20.
DR    Pfam; PF01546; Peptidase_M20; 1.
DR    PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
KW    Complete proteome.
SQ    SEQUENCE   467 AA;  72988D004231F6F7 CRC64;

Alignment Scores:
Pred. No.:
Score:              3.6e-19
Pred. No.:          330.50              Length:     467
Score:              46.21%              Matches:    114
Percent Similarity: 28.79%              Conservative: 69
Best Local Similarity: 11.97%           Mismatches: 166
Query Match:        16                  Indels:     47
DB:                                     Gaps:       17

US-10-014-896-1 (1-1509) x Q9CC46 (1-467)
```

```
Qy    388  CCTGCCCCTGAAGAAGGCTGGAGGTGCCCCATTCTCTCGGGTTGGAGGCTGTGATGGCGGTC  447
           |||                    ||| |||
Db     93  LeuAlaGlyIleAspSerSerArgGlyTyrAlaLeuLeuIleHisGlyHisLeuAspValVal  112

Qy    328  ATCCAAGGCTCGGACCCAGTTGCAGCCGTTCACTTTGATGTGGTG  387
           ||| |||                    ||| |||||||||
Db     93  LeuAlaGlyIleAspSerSerArgGlyTyrAlaLeuLeuIleHisGlyHisLeuAspValVal  112
```

```
Db    113  ProAlaGluThrAlaGluThrTrpSerValHisProPheSerGlyValAlaValGluGlyGln  132
                                    |||            |||  ||||||||||||
Qy    448  ATCTATGGTTCGGGGCACACTGGACACGACAAGAACTCTGTGAGGCATTACTGCAGGGCTTG  507

Db    133  ValTrpGlyTyrArgGlyYalAlaIleAspMetLysAspMetValGlyMetMetIleValAlaAla  152
                                    |||  |||  :::  |||    |||  :::
Qy    508  GAGCTCCTGGTCGATCAGGAGAAGTACATCCCCCGGAAGATCTTTCTTCATTTCCTGGGCCAT  567

Db    153  ArgGlnLeuLysGlnLnAlaLeuGlyIleAlaProProArgAspLeuValPheAlaValAla  172
                                    |||                    :::
Qy    568  GATGAGGAGTCATCAGGGACA---GGGGTCAGAGGATCTCAGCCCTGCTACAGTCAAGG  624

Db    173  AspGluGlnHisGlyYalGlyGlySerTyrArgGlySerGln---TrpLeuValAspAsnArg  189
                                    |||  |||:::  |||
Qy    625  ---------GCGGTCCAGCGTAGCCGTCATCAGTGTGGACCGGAGCAGACA  669

Db    190  ProAspLeuPheAspGlyValThrGluValala---IleGlyGluValGlyGlyProheSerLeu  208
                                    |||  |||                    |||
Qy    670  GATGAGTTCAATTCCT---AACTTCAAGAACGCCATGCCTTGAATGCAGTCTCA  720

Db    209  Thr------ValProCysArgAsnGlyValGlyGluArgGlyLeuTyrLeuIleGluThrAlaAla  226
                                    :::||| ::: |||||||::::
Qy    721  GAGAAGGGGTTCCATGAACCTCATGCTGCCAAGTAAAACATGACTTCAGGCGCCTCTTCAGCT  780

Db    227  GluThrGlyLysMetGlnTrpProMetArgLeuThrAlaArgGlyAlaGlyHisGlySerMet  246
                                    ::: |||                    ||| :::
Qy    781  CCTCCAAAGGAGGACAAGCACCATGGCACTCCTTGCAGCTGTGTCAGCGACGATTGGAGCAGACA  840

Db    247  ValHisAsnGlnAsnAlaValThrAlaValAlaGluAlaGlyAlaGlnValAlaLeuAspValGlyArgHis  266
                                    :::||| ||||||||||||| |||||||||:::
Qy    841  CCAATGCCTATCATATTTGGAAGCGGGACAGTGGTCGCTGTATTGCAGCAACTGGCAAAT  900

Db    267  GlnPheProLeuVal------ThrThrAspThrValValGlnPheLeuAlaLeuAlaIleSerGlu  285
                                    ||||||||||| |||||| :::
Qy    901  GAGTTTCCCTTCCGTCGTCAATATAATATCCTGAGCAACACCATGGCTATTTGAACCACTTATA  960

Db    286  GluThrGlyLeuGlnPheAspIle------GlySerProAspLeu---GluGlyAlaIle  302
                                    ::: |||                    |||
Qy    961  AGCAGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAATAATCAGAAGTTGGGGTGGAGCAGCACTC  1020

Db    303  GluLysGlyGlyProMetAlaArgMetLeuThrAlaArgGlyValAlaGlyLeuLeuTyrAspAsnPro  322
                                    :::  |||                    |||
Qy   1021  ACCATATTCAAAGCAGGGGTCAAGTTCAAATGTCATCCCCCCAGGTGGGCCCAGGCCACAGTC  1080

Db    323  ThrValLeuLysValAlaGlyTyrLysValAlaSnValValProAlaThrAlaGluMetVal  342
                                    |||                    |||
Qy   1081  AACTTCCGGATTCACCCTGGACAG------------ACAGTCCAAGAGGTCCTA  1122

Db    343  AspCysSerArgIleLeuProGlyAlaArgGlnAlaAlaAlaPheGluAlaAlaIleAspProGluLeuIle  362
                                    ::: |||                    |||
Qy   1123  -----GAACTCACGAAGAACATTGTGGTGTGATAACGAGTCCAGTTCACATGTGTAGT  1176

Db    363  GlyProAspValThrArgGluTrpProIleLysAsp-----  373
                                    :::|||||| :::
Qy   1177  GCCTTTGACCCCCTTCAGCCCTTCTGATGACAAGGCCGTTGGGCTACCAGCTGTC  1236

Db    374  ---LeuProProTyrGluThrAlaPheAspGlyAspLeuVal------AspAlaMet  389
                                    ::: |||                    :::
Qy   1237  CGCCAGACGCTACAGTCCGTTCTTCCCGGAAGTCAATATTACTGCCCCAGTTACTTCT  1296

Db    390  AsnAlaAlaValLeuAlaValAlaAspProAsp---GlyArgThrValProTyrMetAlaSer  408
                                    ||| |||                    :::
Qy   1297  GGCAACACAGACGCCGATTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTACCCC  1356

Db    409  GlyGlyThrAspAlaAlaLysPheAlaAlaArgLeuGlyIleAsnTyrSerPheGlyTyrPheThrPro  428
                                    |||||||||||::::                    :::
Qy   1357  ATCTACATACAGCCCTGAA---GACTTCAAACGCATC---CATGGAGTCAACGAGAAAATC  1410

Db    429  LeuArgLeuProThrGlyLeuAspPheThrValAlaLeuPheHisGlyValAlaAspGluArgVal  448
                                    ||| |||||||:::|||:::|||||:::
Qy   1411  TCAGTCCAAGGCCTATGAGACCCAAGTGAAAATTCATCTTTGAGTTGATT  1458
```

```
RESULT 11
Q93H22

ID   Q93H22             PRELIMINARY;      PRT;    441 AA.
AC   Q93H22;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Putative peptidase.
OS   Streptomyces avermitilis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=33903;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=21477403; PubMed=11572948;
RA   Omura S., Ikeda H., Ishikawa J., Hanamoto A., Takahashi C.,
RA   Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osonoe T.,
RA   Kikuchi H., Shiba T., Sakaki Y., Hattori M.;
RT   "Genome sequence of an industrial microorganism Streptomyces
RT   avermitilis: Deducing the ability of producing secondary
RT   metabolites.";
RL   Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).
DR   EMBL; AB070954; BAB69369.1; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR029933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
SQ   SEQUENCE   441 AA;  47835 MW;  277F24D89192681S CRC64;

Alignment Scores:
Pred. No.:
Score:                    4e-18
                          318.00
Percent Similarity:       43.08%
Percent Similarity:       26.79%
Best Local Similarity:    11.518
Query Match:              2

US-10-014-896-1 (1-1509) x Q93H22 (1-441)
```

```
Db       449  SerIleAspAla----------------LeuLysPheGlyIleAspValLeu  460

Qy      661  TTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATCCGCCTTGATTGCAGTCTCA  720
              |||             |||       |||     |||            |||:::
Db      186  Phe----------------SerPheThrValAsnGluLysLeuArgLeuTyrLeuValGlu  202

Qy      721  GAGAGAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCGCAGTCTTCAGT  780
              |||||||     |||         |||   |||   :::      |||:::|||||||||
Db      203  GlnLysGlyMetHisTrpMetArgLeuThrValAspGlyThrAlaGlyHisGlySerMet  222

Qy      781  CCTCCAAAGGAGAACAAAGCATTGGAACCCTTGCACTGCTGTCACCGGATTGGGAGACA  840
              |||||||     |||        |||       |||     |||||||||     |||
Db      223  ThrAsnAspAspAsnAlaIleThrGluLeuCysGluLeuAlaValGlyArgLeuArgHis  242

Qy      841  CCAATGCCTATCATATTGGAAGCGGGACAGCAGGTGGTGACGTGTATTGCGAACTGGCAAT  900
              |||:::     |||||||||        |||||||          |||:::|||||:::
Db      243  ThrTrpProValArgVal---ThrLysThrValArgSerPheLeuAspGluLeuSerAsp  261

Qy      901  GAGTTTCCCTTCCCTCGTCCAATATAATAATCCTGAGCAACCCATGGGCTATTTGAAC...  960
              |||     :::         |||       |||:::      :::  |||:::|||||||
Db      262  AlaLeuGlyThrGluLeu------------------AspProGluAsnMetAspGluThrLeu  276

Qy      961  AGCAGGTTTATGGAGAGAGAAATCCCCTTAACCAATGCGAATAATCAGAACCACCAGCACTC 1020
              |||||||  |||      |||||:::      |||||||:::|||||||:::|||||||
Db      277  AlaLysLeuGlyIleAlaLysMetValGlyAlaThrGluAsnSerAlaAlaPro  296

Qy     1021  ACCATATTCAAAGCAGGGGTCAAGTTGCAAGTGCATCCCCCAGTGGCCCAGGCCACAGTC 1080
              ||||||     |||||||     |||||||   :::         |||  |||||||
Db      297  ThrMetLeuGlyAlaGlyTyrLysValAsnValIlePro...  316

Qy     1081  AACTTCCGGATTCACCCTGGACAGACAGAGTCCAAGAGGTCCTTAGAACTCACGAAG---- 1134
              |||||||||  |||||||  |||  :::|||  |||:::|||  |||||||:::|||
Db      317  AspGlyArgPheLeuProGlyHisGlyGluGlyGluPheLeuGluLeuAlaAspArgIleLeu  336
```

```
RESULT 12
006234

ID   006234             PRELIMINARY;      PRT;    448 AA.
AC   006234;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Hypothetical 48.1 kDa protein (Aminoacylase-1, putative).
DE   DAPE2 OR RV2141C OR MTCY270.27 OR MT2199.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
```

RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37RV;
RX   MEDLINE=98255987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / OSHKOSH;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z95388; CAB08652.1; -.
DR   EMBL; AE007067; AAK46483.1; -.
DR   TIGR; MT2199; -.
DR   TubercuList; Rv2141c; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   448 AA; 48142 MW; 1221321B8A4FB108 CRC64;

Alignment Scores:
Pred. No.:   2.09e-17
Score: 309.50
Percent Similarity: 42.64%   Length: 448
Best Local Similarity: 26.65%   Matches: 125
Query Match: 11.21%   Conservative: 75
                                    Mismatches: 199
                                    Indels: 70
                                    Gaps: 20

US-10-014-896-1 (1-1509) x 006234 (1-448)

RESULT 13
ID   Q9AD91
     Q9AD91   PRELIMINARY;   PRT;   445 AA.

Db   118 yAlaValAspMetLysAspMetValGlyMetMetIleValValValAlaAlaArgHisLeuArgGl 138

Qy   522 CAGGAAGTACATCCCCGACGAGAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTCATC 581
          | ::||| ||| ||||||||| ||||||
Db   138 nAlaIleAlaIleValProProArgAspLeuValPheAlaValAlaAlaAspGluGlnHisGl 158

Qy   582 AGGGACAGGGGGCTCAGAGGATCTCAGCCCCTGCTACACGTCAAGG----------------GG 626
          ||| |||||||||| ||| |||||||||| ||
Db   158 yGlyLysTyrGlySerHis------TrpLeuValAspAsnArgProAspLeuPheAspGl 176

Qy   627 CGTCCAGCTAGCCTTCATTGTGGACGAGGGGGGCTTCATCTTGGATGATTTCATTCCTAA 686
          ||| ||| ::: ||| :::||| ::|||
Db   176 yIleThrGluAla---IleGlyGluValGlyGlyPheSerLeuThr-------ValProAr 193

Qy   687 CTTC----------AAGAGCCCATCCGCCTTGATTGCAGTCTCCAGAGAAGGGTTCATGAA 737
          ::::   ::: ||||| ||||||||||||||| |||||||
Db   193 gHisAspGlyGlyGluArgArgLeuTyrLeuIleGluThrAlaGlyLysGlyIleGlnThr 213

Qy   738 CCTCATCCTGCAAGTAAACACGAGTTCAGCGCCCACTCTTCAGCTCCTTCCAAAGGAGACAAG 797
          ::: ||| ||||||||||| :::||||||| |||||||||||
Db   213 pMetArgGlnThrAlaArgGlyAlaGlyHisGlySerMetValHisAspGlnAsnAl 233

Qy   798 CAATTGGCATCCTTGCAGACCCATGCTGGTCAGCCGATTGGGAGCAGAAGACCAAATGCCTATACTATATT 857
          ||| ||| :::||||| ||||||| ::::   :::||| |||
Db   233 aValThrAlaValCysGluAlaValAlaAlaArgLeuGlyArgHisGlnPheProLeuValCy 253

Qy   858 TGGAAGGGGACAGTGGTGACTGATTGCAGCAACTGGCAAATGAGTTTCCCTTCCTGT 917
          ::: ||||||||||||||| ||||||||||||||| ||||||
Db   253 s---ThrAspThrValAlaGlnPheLeuAlaValValGlyGluGlnThrGlyLeuAlaPh 272

Qy   918 CAATAATAATCCTGACCAAACCCATGGTGTATTTGAACCACACTTATAAGCAGGTTTATGGGAGAG 977
          ::: ||||||||| ||| |||||| :::||||||| |||
Db   272 eAsp----------LeuAspSerProAspLeu----AlaGlyThrIleAspLysLeuGlyProMe 289

Qy   978 AAAATCCCTTAACCAATGCAATAAATCAGGACCACACGGAGACACCAAATGCCTATCATTCAAAGCAGG 1037
          |||||| ::: ||| ::: ||||||||| ||| |||
Db   289 tAlaArgGlyMetLeuLysAspThrAlaAsnProThrMetLeuLysAlaGl 309

Qy   1038 GGTCAAGTTCAATGTCATCCCCCAGTGGCCCCAGGCCACAGTCAACTTCCGGATTCACC 1097
          ::: ||||||||| ||| ||| ||| :::|||||||||||
Db   309 yTyrLysAlaAsnValValProAlaThrArgLeuHisLeuGluThrGlyLeuAlaPh 329

Qy   1098 TGGACAG----------ACAGTCCAAGAGGTCCTA------GAACTCACGAA 1133
          :::|||: ||| ::: ||::   ::|||
Db   329 oGlyArgArgAlaLeuAlaPheGlyLeuAspAlaLeuIleGlyProAspValThrAr 349

Qy   1134 GAACAATTGTGGCTGATAACAGAGTCCAGTGCCATGTGTTGAGTGCCTTTGACCCCTCC 1193
          |||||| :::||||| ||||||||||||
Db   349 tAlaArgMetLeuLysAlaValLeuHisAspThrAlaAsnProThrMetLeuLysAlaGl 369

Qy   1194 CGTCAGCCCTTCTGATGACAAGGCCTTGGGCGTGCCTACCAGCTG----------CTCCGCCGAC 1244
          ||||||| :::||| ||| :::||| ||| ::|||
Db   369 yTyrLysAlaAsnValValProAlaThrArgLeuHisLeuGluThrGlyLeuAlaPh 392

Qy   1245 CGTACAGTCCGTCTCCCGGAAGTCAAATATATTACTGCCCCAGTTACTTCTTATTGGCAAAC 1304
          ||| :::||| ::: ||| ||| |||
Db   373 aValLeuAlaValAlaAspProAsp---GlyArgThrValProTyrMetLeuSerGlyGlyTh 392

Qy   1305 AGACAGCCGAGTTCTTTACAAAACCTCACCACTGGCAATCTACAGGTTCTACCCACTACAT 1364
          |||||| ||||||| ||| ||| ||| :::
Db   392 rAspAlaLysAlaProPheGlyIleGluArgCysPheGlySerProLeuArgIle 412

Qy   1365 ACAGCCTGAA---GACTTCAAAACGCATC---CATGGGAGTCAACGAGAAAAATCTCAGTCCA 1418
          ::: ||: ||| ||| :::||||||| ||| :::|||
Db   412 uProProAspLeuAspPheThrSerLeuPheHisIleGlyValAspGlnArgValProIleAs 432

Qy   1419 AGCC----------TATGAGACCCAAGTG 1437
          ::: |||:::||||||
Db   432 pGlyLeuArgPheGlyThrGluVal 440

Qy   462 CACACTGGACCACAAGAAACTCTGTGATGGGCATTACTGCAGGCCTTGGAGCTTCCTGCTGAT 521
          :::|||| |||::: ||||||| ::::   :::
Db   118 yAlaValAspMetLysAspMetValGlyMetMet... 138

```
AC   Q9AD91;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative peptidase.
GN   SC01676 OR SCI52.18C.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   O'Neil S., Harris D.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RC   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8884336;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2) / M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabbinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).
DR   EMBL; AL590507; CAC36379.1; -.
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   InterPro; IPR005504; RNA_rec_mot.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
DR   PROSITE; PS00759; ARGE_DAPE_CPG2_2; UNKNOWN_1.
DR   PROSITE; PS00030; RRM_RNP_1; UNKNOWN_1.
SQ   SEQUENCE   445 AA;  47879 MW;  95C6E4F7E2D56CE6 CRC64;

Alignment Scores:
Pred. No.:              1
Score:                  1.32e-16        Length:          445
Percent Similarity:     300.00          Matches:         121
Best Local Similarity:  44.2%           Conservative:    69
Query Match:            28.2%           Mismatches:      181
DB:                     10.86%          Indels:          58
                        16              Gaps:            18

US-10-014-896-1 (1-1509) x Q9AD91 (1-445)
```

```
Qy   79 GlyGluAspProSerArgProAlaLeuLeuIleHisGlyHisThrAspValValProAla 98
Db  454 GGTCGGGGCGACACTGGACGACAAGAAACTCTGATGAGGCATTACTGCAGGCCTTGGAGCTC 513
        |||||||||  :::|||   |||:::        |||   ||| :::|||:::|||
Qy  394 CCTGAAGAGACGGCTGGGAGGTGCCCCATTTCTCTGGGTTGGAGGGTGAGGGTCATCTAT 453
Db   99 AsnAlaAspAspTrpThrHisHisProPheSerGlyLeuValAlaAspHisHis 118
Qy  514 CTGCTGCATCAGGAAGTACATC---- 558
Db  119 GlyArgGlyAlaValAlaAspMetLysAsp 133
```

(Additional DNA/protein alignment blocks continue across the page in the same Qy/Db paired format, with query nucleotide positions and database protein positions; see original image for full sequence detail.)

```
Qy 334 GGCTCGGACCCCAGCTTGCACCCTACCTGCTCACTTTGATGGCTGTGCTGCC 393
       |||  ||||||||||||||||||
Db 414 AspPheHisGlyMetPheHisGlyValAlaAspGluArgValProValGluGlyLeuGlnPhe 433
```

```
Qy      1432 CAAGTGAAATTCATCTTTGAGTTGATT 1458
             ||||||                  |||
Db       434 GlyValArgValLeuAspArgPheIle 442

RESULT 14

ID   Q93R29
AC   Q93R29;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE   Putative peptidase
GN   SCO571 OR SC5G5.03c.
OS   Streptomyces coelicolor.
OC   Bacteria; Firmicutes; Actinobacteria;
OC   Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX   NCBI_TaxID=1902;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Collins M.R., Harris D.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RA   Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2);
RX   MEDLINE=97000351; PubMed=8843436;
RA   Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA   Kinashi H., Hopwood D.A.;
RT   "A set of ordered cosmids and a detailed genetic and physical map for
RT   the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL   Mol. Microbiol. 21:77-96(1996).
RN   [4]
RP   SEQUENCE FROM N.A.
RC   STRAIN=A3(2); /M145;
RA   Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
RA   Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H.,
RA   Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
RA   Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
RA   Huang C.-H., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S.,
RA   Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,
RA   Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
RA   Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J.,
RA   Hopwood D.A.;
RT   "Complete genome sequence of the model actinomycete Streptomyces
RT   coelicolor A3(2).";
RL   Nature 417:141-147(2002).;
EMBL;  AL591823; CAC39633.1;
DR   InterPro; IPR001261; ARGE_DAPE_CPG2.
DR   InterPro; IPR002933; Peptidase_M20.
DR   Pfam; PF01546; Peptidase_M20; 1.
DR   PROSITE; PS00758; ARGE_DAPE_CPG2_1; UNKNOWN_1.
SQ   SEQUENCE    442 AA;  36762DD638C2CA28 CRC64;


Db      434 GlyValArgValLeuAspArgPheIle 442

Alignment Scores:
Pred. No.:
Score:                1.39e-14
Percent Similarity:   40.60%
Best Local Similarity: 23.20%
Query Match:          9.99%
DB:

Length:      442
Matches:     116
Conservative: 87
Mismatches:   167
Indels:      130
Gaps:        19

US-10-014-896-1 (1-1509) x Q93R29 (1-442)
```

```
Qy       148 GCGATGAAAAGAGGCCCTGAAAAGGTGCCATCCAGATTCCAACACAGTGACTTTTAGCTCGAG 207
             |||       :::|||||||  ||||||| :||||     ||
Db        11 AlaSerGluGluAlaGlnArgGluValValAspLeuCysAlaGluLeuIleArgPheAsp 30
```

```
Qy       208 AAGTCCAATACTACAGCC------------------------------------CTGGCT 231
             ||||||               |||:::
Db        31 ThrSerAsnProThrSerAspGlyArgGlyAlaCysAlaAlaAspTrpValAlaAlaGluLeuAla 50
```

```
Qy       232 GAGTTCGGAAAATACATTCATTCATAAAGTTCTTTCCTACAGTGGTCACCAGCTTTATCCAG 291
             |||      :::||||        ||:::||||||||   |||:::
Db        51 GluAlaGly------------------------------------IleAla 55
```

```
Qy       292 CATGAAGTCGTGGAA------------------------------GAGTATAGCCACCTGTTCACTATCCAAGGCTCG 339
             |||                    :::|||||||
Db        56 SerGluLeuValGluSerAlaProGlyArgAlaAsnValAlaValAlaArgIleProGlyAla 75
```

```
Qy       340 GACCCCAGCTTGCAGCCCCTACCTGCTGATGGTCACTTTGATGTTGGTGGGCCCCCTGAA 399
             |||           :::|||||||     |||||||||
Db        76 AspThrSerArgGlyAlaLeuLeuValHisGlyHisLeuAspValValProGlyAlaSerAla 95
```

```
Qy       400 GAAGGCTGGGAGGTGCCCCCATTCTCCTGGCGTTGGAGCGGTGAATGCGGTCATCTATAGGTCGG 459
             |||||||||||   :::|||||||||         :::|||||||
Db        96 AlaGluTrpArgValProProPheSerGlyGlyLeuGlyAspGlyTyrLeuThrTrpLysArg 115
```

```
Qy       460 GGCACACTGGACGACGACAAGAACTCTGTGATGGCATTACTGGAGGCCTTGGAGCTCCTGCTG 519
             |||||||||||       :::|||||||
Db       116 GlyAlaIleLeuAspMetLysAspThrValAlaValMetLeuAlaThrAlaArgHisPheAla 135
```

```
Qy       520 ATCAGGAAGTACATCCCCGGAAGATCTTTCTTCATTCTGTGGGCCATGATGAGGAGTCA 579
             |||||||        :::||||||
Db       136 ArgThrGlyThrArgProAlaArgGlyIleLeuValLeuAlaPheLeuAlaAspGluGluAla 155
```

```
Qy       580 TCAGGGACA---GGGGCTCAGAGGATCTCACCCTGCTCACAGTCGAAGGGGCGTCCAGCTA 636
             |||||||        :::|||||||
Db       156 GlyLeuLysPheGlyAlaHis------------TrpLeuValGluHisArgProGluLeuPhe 172
```

```
Qy       637 GCCTTCATT---------------GTGGACGAGGGGGGGC------TTCATCTTGGATGATTTC 678
             ||||||              :::||||||||        |||||||
Db       173 AlaGlyValThrGluAlaIleGlyGluGlyGlyGlyPheSerTyrArgAlaLeuAspAspAspThr 192
```

```
Qy       679 -----------------ATTCCTAACTTCAAGAAGAGCCCCTTGATTGCAGTCTCAGAG 723
             |||                :::|||        :::|||
Db       193 ArgArgLeuTyrProIleGluAsnAlaGlnArgGlyMetAlaArgPheMetGluLeuThrAla 212
```

```
Qy       724 AAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGGCCACTCTTCAGCTCCT 783
             |||    :::|||||||
Db       213 AlaGly------------ArgAlaGlyHisGlySerSerPro 222
```

```
Qy       784 CCAAAGGAGAACAAGCATTGGCCATCCTCCGAGCTGCTGCTGCGACGCCGATTGGAGCAGCAGACACCA 843
             |||||||||   |||:::|||||||
Db       223 AsnAspGluAsnAlaValThrAspLeuAlaGluSerLeuThrArgIleGlyArgHisThr 242
```

```
Qy       844 ATGCCTATACATATATTGGAGACGGGGACAGTGGTGACTGTTGCAGCAAACTGGCAAAATGAG 903
             |||                :::|||
Db       243 PhePro------------ValArgLeuIleGlu 249
```

```
Qy       904 TTTCCCTTCCCTGTCAATATAATCCTGACCAACCCATGGCTATTTGAACCACTTT------ 957
             |||       :::|||||||
Db       250 ------ProValArgAlaLeuLeuLeuAlaGluLeuAlaAlaArgLeuGlnAlaValAspLeu 266
```

```
Qy       958 -----------------------------ATAAGCAGGTTT 969
             |||                             :::|||
Db       267 AspLeuAspAlaGlyAspLeuGlnLeuAlaLysLeuGlyHisValAlaAspPhe 286
```

```
Qy       970 ATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACACGGCGACTCACCAATATTC 1029
             |||||||:::|||||||
Db       287 MetGln------------------ValValLeuArgAsnSerAlaAsnProThrMetPhe 300
```

```
Qy      1030 AAACGAGGGGTCAAGTTCAAATGTCATCCCCCCAGTGGGCCCAGCCCACAGTCAACTTCCGG 1089
             |||||||     :::|||||||
Db       301 ThrAlaGlyTyrGlnThrAsnValIleProGlyArgAlaThrAlaArgValAspGlyArg 320
```

```
Qy      1090 ATTCACCCTGGAGCAGACAGTCCAGAGAACTCACGAGAGAACAATTGTGGGCTGAT 1149
             |||||||      :::|||:::|||
Db       321 PheLeuProGlyHisGlu---GlnGluLeuIleAsp------------------------ 331
```

```
Qy      1150 AACAGAGTCCAGTTCCATGTGTGAGTGCTTTGACCCCCTCCCCGTCAGCCCTTCGAT 1209
```

Db 332 --------------------------------ThrIleAspAlaLeuLeuProSerValSer 342
                                    ||| :::
Qy 1210 GACAAGGCCTTGGGCTACCAGCTGCCCGGCCAGACC---------------------------- 1245
        ::: ||| :: ||
Db 343 ArgGluTrpValAsnHisAspIleAlaMetGluThrSerPheAspGlyProLeuValAsp 362
        ::: :|| :::||
Qy 1246 ------GTACAGTCCGTCTTCCCGGAAGTCAATATTACTCCCCAGTTACT 1290
        |||||||
Db 363 AlaMetCysAlaAlaValArgAlaGluAspProAspGlyHis---ProValProTyrCys 381
        ||| :::
Qy 1291 TCTATTGGAACAACAGAGACGACAGCCGAATTCTTTACAAACCTCACCACCTGGAACACTTTC 1350
        ||| ||||||||
Db 382 AsnProGlyGlyThrAspAlaLysAlaPheThrLysLeuAsnSerPheGlyPhe 401
        ::: :||
Qy 1351 TACCCCATCTACATACAGCCCTGAA---GACTTCAAACGCATC---CATGGAGTCAACGAG 1404
        :::||| :||||| |||||||::||
Db 402 LysGlyLeuLysProHisAspTyrGlyArgLeuPheHisGlyValAspGlu 421
        ::: |||||::|| :::|||
Qy 1405 AAAATCTCAGTCCAAGCCTATGAGACCCAAGTGAAAATTCAATCTTTGAGTTGATTCAGAAT 1464
        ||||||| ||||:||
Db 422 ArgValProLeuGluGlyLeuArgPheGlyValArgValMetThrArgLeuTrpGlnAsp 441

RESULT 15
Q9VCO9
ID Q9VCO9   PRELIMINARY;   PRT;   401 AA.
AC Q9VCO9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CG6733 protein.
GN CG6733.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Murphy D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287;2185-2195(2000).
DR EMBL; AE003741; AAF56097.1; -.
DR FlyBase; FBgn0039052; CG6733.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
SQ SEQUENCE 401 AA; 44925 MW; 97E6AE6568EDF94E CRC64;

Alignment Scores:
Pred. No.:   1.06e-12
Score: 253.50
Percent Similarity: 38.77%   Matches: 116
Best Local Similarity: 24.58% Conservative: 67
Query Match: 9.18%   Mismatches: 180
                               Indels: 109
                  5   Gaps: 19

US-10-014-896-1 (1-1509) x Q9VCO9 (1-401)

Qy 130 AGCAAAAGAGGAACGGCGTCGCGATGAAGAAGAGGCGGCTGAAAAGGTGCCATCCAACA 189
        ::: |||||| ||||||| :::||| ||||||| ||||||:::
Db 8 AsnAsnGluGluIleLysIlePheArgGluTyrLeuArg--------------IleProThr 23

Qy 190 GTGACTTTTAGCTCTGAGAGAGTCCAATACTACAGCCCCGGCTTGGCTTGCTGTTCGGAAAATAC--- 246
        ||| :::||| :::
Db 24 Val------HisProAspValAspTyrThrAlaCysValGluPheLeuLysArgGln 40

Qy 247 ------ATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTT 285
        ||||:||:||| ||| :::
Db 41 AlaSerSerLeuAsnLeuProValGluValValTyrProAlaValGlnThrLysProVal 60

Qy 286 ATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCTGTTCACTACTATCCAAGGCTCCGGACCC 345
        ||||| |||||||::||||
Db 61 Val----------------IleIleLeuLysTrpGluGlySerGlnPro 70

Qy 346 AGCTTGCAGCCCTACCTGCTCGATGAGTGGTCACTTTGATGTGGTGCCTGCCTGCCCCTGAAGAAGGC 405
        :::|| :::|||
Db 71 GluLeuSerSerIleLeuValLeuAsnSerHisThrAspValProValProValPheArgGluLys 90

Qy 406 TGGGAGGTGCCCCCAATTCTCTGGG---TTGGAGCCTGAATGGCGTCATCTRATGGTCGGGGGC 462
        ||| :::
Db 91 TrpThrHisGlyProPheSerAlaAspIleAspGluGluGlyGlyArgIlePheAlaArgGly 110

Qy 463 ACACTGGACGACAAGAACTCTGTGATGGACATTACTGCAGGGCCTTGGAGCCTCCTGTGATC 522
        ||| ||| |||:||| |||||||
Db 111 ThrGlnAspMetLysSerValGlyThrGlnThrGlnTyrLeuGlnAlaIleArgLeuLeuLysAla 130

Qy 523 AGGAAGTAACACCCCGGAAGAGAATCTTTCTTCATTTCTCTGGGCCCATGATGAGGAGTCATCA 582
        ::: ||||||
Db 131 SerGlyPheLysProLysArgAsnLeuTyrValThrProPheValProAspGluGluThrGly 150

Qy 583 GGG---ACAGGGCGCAGAGGATCTCAGCCCTGCTACAGTCAAGGGCGGGCCAGCTAGCC 639
        |||
Db 151 GlyHisLeuGlyMetLeuAlaGluPheValLysThrAspTyrTyrLysLysMetAsnAlaGly 170

Qy 640 TTCATGTGGTGGACGAGGGGGGCTTCATCCTGGATGAGTTCAATTCCTAACTTCAAGAGCC 699
        ||| :::
Db 171 PheSerLeuAspGluGlyAlaThrSerGluSerAsp------------------182

Qy 700 ATCGCCCTTGATTCGACTGTCAGTAGAAAGGGGTTCCATGAAACCTCATGGCTGGAAGTAAACATG 759
        ::: |||||||::||:::
Db 183 ValHisHisLeuPheTyrArgLeuGlyArgGluArgTrpGlyLeuLysLeuLysValSerGly 202

Qy 760 ACTTCAGGCCACTCTTCAGCTCCTCCAAAGGAGACAAGCAGCATTGGCATCCTTGCACTGGCT 819
        ||||||||||| |||
Db 203 ThrSerGlyHisGlySerLeuLeuLeuProLeuAsnThrAlaGlyValLysLeuAsnTyrLeu 222

Qy 820 GTCAGCCGGATTGGAGGCAGACACCAATGCCTATCATATTTTGGAAGGGGGACAGTGGTGTGACT 879
        ||| |||
Db 223 ValAsnLysLeuThrGlu------------------------PheArgThrSerGlnVal 234

```
Qy    880  GTATTGCAGCAACTGGCAAAATGAGTTTCCCTTCCCGTTGTCAAATATAATCCTGAGCAACCCA  939
                       ||||||  :::
Db    235  ------------GluAsnLeuAlaArgAsp-----                                  240

Qy    940  TGGCTAATTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATA  999
                       :::  ||||||:
Db    241  ------------------------SerSerLeuSerLysGlyAsp  247

Qy   1000  ATCAAGGACCACCACGGCCACTCACCATATTCAAAGCAGGGGGTCAAGTTCAATGTCATCCCC  1059
                       ||:||||  ||||||                        ||||||||:::  |||||||:::||
Db    248  Val---ThrThrValAsnLeuThrGlnLeuSerGlyValGlnSerAsnValValPro  266

Qy   1060  CCAGTGGCCCAGGCCACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTC  1119
                       ||:||:  :::||:||  :::  ||||||
Db    267  ProLeuPheGluAlaValPheAspIleArgIle-----  277

Qy   1120  CTAGAACTCACGAAGAACATTGTGGCTGATAACAGA----------------------------  1155
                       ::::||  :::  |||
Db    278  ---AlaIleThrValAsnValValAlaPheGluLysGlnIleArgAspTrpCysGluGlu  296

Qy   1156  ------------------------GTCCAGTTCCATGTGTTGAGTGCCTTTGACCCCTCCCC---  1194
                                             ::  |||                               |||
Db    297  AlaGlyGlyGlyIleGluIleAspPhe------------------PheGlnLysGluProTyr  311

Qy   1195  GTCCAGCCCTTCTGAT---GACAAGGCCTTGGGCTACCAGCTGTCCGCCAGACCGTACAG  1251
                       |||||||||  |||  :::                                  :::||:::
Db    312  IleGlyProThrLysLeuAspAsnSerAsnProTyrTrpLeu--------AlaValLys  328

Qy   1252  TCCGTCTCCCGGAAGTCAATAATTACTGCC---CCAGTTACTTCTATTGGCAACACAGAC  1308
                       |:||:||  :::  :::                :::  ||||||
Db    329  AlaAlaIleAspGluLeuGlyLeuValHisProIleValCysProGlyAlaThrAsp  348

Qy   1309  AGCCGATTCTTTACAAACCTCACCACTGGCCATCACAGGTTCTACCCATCTACATACAG  1368
                       ||||||||  :::                                  |||  |||||
Db    349  SerArgPheIleArgGluLysGlyThrProAlaIleGlyPheSerProIleIle----  366

Qy   1369  CCTGAAGACTTCAAACGCCATCCATGGAGTCAACGAGAAAATCTCAGTCCAAGCCTATGAG  1428
                       ||||||||||  |||  :::                     :::  |||  |||  :::
Db    367  ---AsnThrThrMetArgIleHisAspGluPheLeuGlnAlaAspValTyrLeu  385

Qy   1429  ACCCAAGTGAAATTCAACTTTGAGTTGATTCAGAAT  1464
                       :::  :::::::||:::||:::|||
Db    386  AsnGlyIleAspValTyrLysLysIleIleArgAsn  397
```

OM nucleic - protein search, using frame_plus_n2p model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: June 27, 2003, 10:54:48 ; Search time 24 Seconds
                 3699.932 Million cell updates/sec
                 (without alignments)

Title: US-10-014-896-1
Perfect score: 2762
Sequence: 1 atggctcaagcggtgcgtttg........ctcacctgcacaaactgtga 1509

Scoring table: BLOSUM62
                 Xgapop 10.0 ; Xgapext 0.5
                 Ygapop 10.0 ; Ygapext 0.5
                 Fgapop 6.0 ; Fgapext 7.0
                 Delop   6.0 ; Delext   7.0

Searched: 262574 seqs, 29422922 residues
                 Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%

Command line parameters:
-MODEL-frame+_n2p.model -DEV-xlp
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104415_10416/app_query.fasta_1.1671
-O=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104415_10416/app_query.fasta_1.1671
-DB-Issued_Patents_AA.* -QFMT=fastan -SUFFIX-n2p.rai -MINMATCH-0.1 -LOOPCLL-0
-LOOPEXT-0 -UNITS-bits -START=1 -END=-1 -MATRIX-blosum62 -TRANS-human40.cdi
-LIST-45 -DOCALIGN-200 -THR_SCORE=pct -THR_MAX-100 -THR_MIN-0 -ALIGN-15
-MODE-LOCAL -OUTFMT-pto -NORM=ext -HEAPSIZE-500 -MINLEN-0 -MAXLEN-200000000
-USER=US10014896_@CGN_1_1_28_@runat_27062003_104415_10416 -NCPU-6 -ICPU-3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK-100 -LONGLOG
-DEV_TIMEOUT-120 -WARN_TIMEOUT=30 -THREADS-1 -XGAPOP-10 -XGAPEXT-0.5 -FGAPOP-6
-FGAPEXT-7 -YGAPOP-10 -YGAPEXT-0.5 -DELOP-6 -DELEXT-7

Database : Issued_Patents_AA.*
   1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
   2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
   3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
   4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
   5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep:*
   6: /cgn2_6/ptodata/1/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                              SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	201	7.3	418	4	US-09-134-001C-3552	Sequence 3552, Ap
2	190	6.9	373	4	US-09-814-951A-2	Sequence 2, Appli
3	185.5	6.7	408	1	US-08-127-278-4	Sequence 4, Appli
4	185.5	6.7	408	4	US-08-555-860-4	Sequence 4, Appli
5	185.5	6.1	408	4	US-09-814-951A-4	Sequence 4, Appli
6	168	6.1	431	4	US-09-134-001C-4338	Sequence 4338, Ap
7	154.5	5.6	446	4	US-09-134-001C-5125	Sequence 5125, Ap
8	138.5	5.0	383	4	US-09-575-602-4	Sequence 4, Appli
9	118.5	4.3	389	4	US-09-134-001C-3161	Sequence 3161, Ap
10	116.5	4.2	412	4	US-09-285-055-2	Sequence 2, Appli
11	114.5	4.1	430	4	US-09-134-001C-4302	Sequence 4302, Ap
12	113	4.1	1284	4	US-09-343-494-9	Sequence 9, Appli
c 13	108	4.0	1253	3	US-08-479-722B-4	Sequence 4, Appli
c 14	107.5	4.0	1251	5	PCT-US95-02251-3	Sequence 3, Appli
c 15	107.5	4.0	1252	1	US-08-199-780-3	Sequence 3, Appli
c 16	107.5	4.0	1252	4	US-08-316-650-3	Sequence 3, Appli
17	107	3.9	402	4	US-09-319-892-4	Sequence 4, Appli
18	105	3.8	642	1	US-08-706-936-5	Sequence 5, Appli
19	102	3.7	3969	4	US-08-061-376-5	Sequence 5, Appli
20	101	3.7	412	4	US-09-423-439-10	Sequence 10, Appl
21	101	3.7	642	4	US-09-423-439-16	Sequence 16, Appl
22	101	3.7	643	4	US-09-423-439-26	Sequence 26, Appl
23	101	3.7	666	4	US-09-423-439-51	Sequence 51, Appl
24	101	3.7	673	4	US-09-423-439-32	Sequence 32, Appl
25	100.5	3.6	1224	4	US-09-930-872-4	Sequence 4, Appli
26	100	3.6	862	1	US-08-325-267A-4	Sequence 22, Appl
27	99.5	3.6	750	4	US-09-165-239A-4	Sequence 4, Appli
28	99.5	3.6	894	3	US-08-362-520-22	Sequence 22, Appl
29	99.5	3.6	894	1	US-08-971-692-15	Sequence 15, Appl
30	99.5	3.6	1481	2	US-08-616-844-40	Sequence 40, Appl
31	99.5	3.6	1481	2	US-08-599-654-40	Sequence 40, Appl
32	99.5	3.6	1481	3	US-08-944-868A-40	Sequence 40, Appl
33	99.5	3.6	1481	3	US-08-944-423A-40	Sequence 40, Appl
34	99.5	3.6	1481	3	US-08-944-496-40	Sequence 40, Appl
35	99.5	3.6	1537	1	US-08-325-267A-2	Sequence 2, Appl
36	99	3.6	1610	4	US-09-513-783A-22	Sequence 22, Appl
37	99	3.6	1125	4	US-09-513-783A-152	Sequence 152, Ap
38	98.5	3.6	446	1	US-08-485-216-2	Sequence 2, Appli
39	98.5	3.6	446	2	US-09-003-245-2	Sequence 2, Appli
40	98.5	3.6	446	4	US-08-853-552-2	Sequence 2, Appli
41	98	3.5	647	4	US-09-423-439-60	Sequence 60, Appl
42	97.5	3.5	1804	4	US-09-362-336A-4	Sequence 4, Appli
43	96.5	3.5	2627	2	US-09-751-186A-4	Sequence 4, Appli
44	96.5	3.5	2627	2	US-09-060-836-3	Sequence 3, Appli
45	96.5	3.5	2627	4	US-09-184-445-3	Sequence 3, Appli

RESULT 1
US-09-134-001C-3552
Sequence 3552, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3552
LENGTH: 418
TYPE: PRT
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3552

                              ALIGNMENTS

US-09-134-001C-3552

Alignment Scores:
Pred. No.:
Score: 201.00   Length: 418
Percent Similarity: 38.66%   Matches: 106
Best Local Similarity: 22.89%   Conservative: 73
Query Match: 7.28%   Mismatches: 196
                                         Indels: 88
                                         Gaps: 19

US-10-014-896-1 (1-1509) x US-09-134-001C-3552 (1-418)

Qy 178 CAGATTCCAACACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTGCAGTTC 237
        ::::::   ::   ::::   :: :::::   ::   ::   :::::   :: ::   :::::::
Db   1 LysValLysIleIleThrLysAsnGluHisAspCysGluMetThrValLeuSerGluGln 20

```
Qy   238  GGAAAATACATTCATAAAGTCTTTCCTACAGTGGTCAGCACCAGCTTTATCCAGCATGAA  297
               |||   :::      ||||||
Db    21  AspLys---IleArgLeuLeuAlaAspIleValLysIleGlnThrGluAsnIleGlu     39

Qy   298  GTC---GTGGAAGAGTAT                                            321
          |||    ::: ||||||
Db    40  IleGluValCysGluThrLeuLysAspLeuLeuSerGlnThrTyrAspIleAspSerIle  59

Qy   322  TTCACATACATCCAAAGGCCTCGGACCCCAGCTTG---                        360
                                            ---CAGCCCTAC
Db    60  ValLysValAlaAsnAspSerArgAlaAsnLeuValAlaGluIleGlySerGlyAlaProVal 79

Qy   361  CTG---CTGATGGCTCACTTTGATGTGGCTGCCCCTGAA---GAAGGCTGGGAGGTG     414
          |||   |||||||||||||||||||||||||||
Db    80  LeuAlaIleSerGlyHisMetAspValValAlaGlyAspHisAspTrpThrPhe        99

Qy   415  CCCCCAATTCTCTGGGTTGGACGGTGAGGGTCACAACTGGACGAC                474
          ||||||||||||||||||| ::: |||||||||||||||
Db   100  ProProPheGluLeuThrAspLysAspGlyLysLeuPheGlyLysGlyThrThrAspMet 119

Qy   475  AAGAACTGTGGTGATGCCATTACTGCAGGCCTTGGAGCTCCTGCTGATCAGAAGTACATC  534
          |||   |||||||||||||||||||||||||
Db   120  LysGlyGlyLeuMetAlaMetValIleLeuGluLeuCysGlnSerAsnAlaLeu        139

Qy   535  CCCCGAAGAGATCGTTTCTTCATTTCTCTGGGCCATGATGAGGAAGAGGGACAGGGGCT   594
          |||||||||||||||||||||||
Db   140  LysGlnGlyThrIleArgLeuLeuAlaThrThrGlyGluGluThrGlnThrGlyAla     159

Qy   595  CAGAGGAATCTCAGCCCTGCTACAGTCCAAGGGGCGGTCCAGCTGGCCTTCATTGTGGGAGCGAG 654
          |||   :::
Db   160  Gln------LeuLeuAlaAspGlyGly                                   171

Qy   655  GGGGGCTTCAPCTTGATGANTTCANTCCNTAACTTCAAGAGAAGCCCATGGCCGTTGAGTTGCA 714
          |||   :::                  TyrLeuAspAspVal
Db   172  SerGlyLeuIleIleIleGlyGlu------                                184

Qy   715  ---GTGTCAGAGAAGGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCAC  771
             |||||||||||||||||||||||:::|||
Db   185  TyrTyrValAlaHisLysGlySerMetSerCysTyrValValThrAlaLysGlyLysAlaAlaHis 204

Qy   772  TCTTCAGCTCCTCCAAAGGAGGACAAGCACATTGGCATCCTTGCAGCTGCTGTCAGCCGATTG 831
               ||| :::
Db   205  SerSerMetProHisLeuGlyThrAsn                                   213

Qy   832  GAGCAGACACCAAATGCCTATCATATTTTGGAAGCGGGACACTGGTGACTGTGATTGCAGCAA 891
          |||   :::|||
Db   214  ---AlaValAspIleLeuLeuValAsp                                   220

Qy   892  CTGGCAAATGAGTTCCTGTCCTCTGCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAA 951
          |||||||     :::
Db   221  PheValAsnGluMetLysGlnGluThrTyrLysAsnIleLysGluHisAspLysLysValHisGlu 240

Qy   952  ------CCACTTATAAGCAGGTTTATG                                   981
                |||   :::                      GAGAGAAAT
Db   241  LeuAspAlaValProMetIleGluLysLysHisIleHisArgHisValIleGlyGluSer  260

Qy   982  CCCTTAACCAAATGCAATAATCAGGACCACACCAGCCACTCACCATTCAAAGCAGGGGTC  1041
          |||   :::                     ::::|:|||
Db   261  HisIleTyrSerGlyPheValMetLeuAsn------SerValPheAsnGlyGlyLys     277

Qy  1042  AAGTTCAAATGTCAATCCCCCAGTGGCCCAGCCCAACTTCCGGGATTCACCCT---      1098
          |||||||                    :::
Db   278  GlnValAsnSerValProHisLysIleTyrAsnValArgThrValProGlu          297

Qy  1099  ------GGACAGAGTCCAAGAGAGTCTAGAACTCACGAGAACATGTGGGCTGATAAC     1152
                |||   :::
Db   298  TyrAspSerThrPheValLysAspLeuGluLysValIleArgHisValGlyGlyAsp     317

Qy  1153  AGAGTCCAGTCCAAGTGTTGAGTGGCCTTTGACCCCGTCAGCCCTTCTGATGAC        1212
          |||   :::
Db   318  TyrLeuThrValAspIleProSerSerHisAsp---                          335

Qy  1213  AAGGCCTTGGGCTACCAGCTGCTCCGCAGACCGTACAGTCCGTCTTCCGGAAGTCAAT    1272
          |||   :::||||||
Db   336  AsnProLeuIleGlnAsnIleThrArgIleAlaAlaProAsnTyrValHisAspIleVal  355

Qy  1273  ATTACTGCCCCAGTTACTTCTTGATGGCCAACACACAGCACAGCACGGATTCTTTACAAACCTCACC 1332
          :::::||                 ------LeuIleGlyThrHisArgAspAlaSerSerPheLeuGlyThrAsn
Db   356  ValSerAla------                                               372

Qy  1333  ACTGGCGATCTACAGGTTCTACCCCATCTACATACAGCGTACTGAAGACTTCAAACGGATCCAT 1392
          |||                     ------AlaValPheGlyProGlyGluSerIleMetAlaHis
Db   373  GluAsnAsnValAspPhe------                                      390

Qy  1393  GGAGTCAACGAGAAAATCTCAGTCCAAAGCCTATGAGACCCAAAGTG------AAATTC   1443
          |||   :::                                        ---  |||
Db   391  GlnValAspGluPheIleArgLysAspMetTyrLeuSerThrTyrIleLeuAspSerTyrIleLysSerAsp 410

Qy  1444  AICTTTGAG                                                     1452
          :::|||:::
Db   411  ValPheLys                                                     413
```

RESULT 2
US-09-814-951A-2
Sequence 2, Application US/09814951A
Patent No. 6387661
GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE, AND USES THEREOF
; FILE REFERENCE: CL001179
; CURRENT APPLICATION NUMBER: US/09/814,951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-951A-2

Alignment Scores:

Pred. No.:	
Score:	1.76e-10
Percent Similarity:	190.00
Best Local Similarity:	39.27%
Query Match:	23.82%
	6.88%
	4

Length:	373
Matches:	91
Conservative:	59
Mismatches:	158
Indels:	74
Gaps:	14

US-10-014-896-1 (1-1509) x US-09-814-951A-2 (1-373)

```
Qy   307  GAGTATAGCCACCTGTTCACTATCCAA------GGCTCGGACCCCAGCTTG          351
          :::::|||                   ------  |||:::::|:||||
Db    18  GlnTyrLeuArgIleArgThrValGlnProLysAspAspTyrGlyThrThrAsnProThrLeu 37

Qy   352  CAGCCCTACCTGCTCGATGGCTCACTTTGATGTGGCTGCCCCTGAACAAAAGGCTGGGAG  411
          ||||||                  ::|||
Db    38  SerSerIleLeuGlnLeuAsnSerHisThrAspSerValValPheLysValGluHisTrpSer 57

Qy   412  GTGCCCCAATTCTCTGGGTTG---GAGGGTGATGGCGGTCATCTCATGGTCGGGGGCCACTG 468
          |||   :::            ---
Db    58  HisAspProPheGluAlaPheLysAspSerGluGlyTyrIleTyrIleArgGlyAlaGln    77

Qy   469  GACGACAAGAACTCTGTGATGGCAATTACTGCAGGCCGTTGGAGCTCCTGCTGATCAGGAG  528
          ||||||                    :::
Db    78  AspMetLysCysValSerGlyIleGlnTyrLeuGlnAlaValAlaArgArgLeuGlyGly   97

Qy   529  TACATCCCCGAAGATCTTTCTTCATTTCTCTGGGCCATGATGAGGACTGTCATCAGGGACA  588
          :::|||
Db    98  HisArgPheProAlaThrIleHisMetThrProPheValProAspGluValAlaGlyGly--- 116

Qy   589  GGGGCTCAGAGGAGTCTCAGCCCTGCTACAGTCCGTCCAGCTA------             636
          ||||||                                    ------
Db   117  ------HisGlnIleGlyMetGluLeuPheValGlnArg                       126

Qy   637  ---GCCTTCATTGTGGGACGAGGGGGCTTCATCTTGATGAATTCATTCCTAACTTCAAG   693
```

RESULT 3
US-08-127-278-4
Sequence 4, Application US/08127278
Patent No. 5498697
GENERAL INFORMATION:
; APPLICANT: IWAKI, Kanso
; APPLICANT: OHTA, Tsunetaka
; APPLICANT: KUROTO, Masahi
; TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
; TITLE OF INVENTION: PREPARATION OF SAID PROTEIN
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BROWDY AND NEIMARK
; STREET: 419 Seventh Street, N.W., Suite 300
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20004

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/127,278
; FILING DATE: 27-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 281136/1992
; FILING DATE: 28-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: NEIMARK, Sheridan
; REGISTRATION NUMBER: 20,520
; REFERENCE/DOCKET NUMBER: IWAKI-2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; TELEX: 248633
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 408 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-127-278-4

Alignment Scores:
Pred. No.:
Score: 5.27e-10   Length: 408
Percent Similarity: 185.50   Matches: 89
Best Local Similarity: 39.14%   Conservative: 57
Query Match: 23.86%   Mismatches: 158
                          6.72%   Indels: 69
                          1   Gaps: 13

US-10-014-896-1 (1-1509) x US-08-127-278-4 (1-408)

```
Qy  319 CTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTT 378
            |||   ||||||:::||||||||:::||:||:|||  |||||| ::||||
Db   62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLeuAsnSerHisThr 81

Qy  379 GATGTGGTGCCTCGCCCCTGAACAAGGCTGGGAGGTGCCCCCCATTCTCTGGGTTG--GAG 435
            ::|    ||||||||||||:::||||||||:||||||||||
Db   82 AspValValProValPheLysGluHisTrpSerHisAspProHisGluAlaPheLysAsp 101

Qy  436 CGTGATGGCGTCAATCATATGGTCGGGGGCCACACTGGACAAGAACTCTGATGGCATTA 495
            :::||||||||||:::||||||:::||||||||||||||||
Db  102 SerGlyGlyTyrIleTyrAlaArgGlyAlaGlnAspMetLysCysValSerIleGlnTyr 121

Qy  496 CTGCAGGCCTTGGAGCTCCTGCTGATCAGGAGAGTACAATCCCCGAGAGAATCTTTCTCATT 555
            |||||||||||||||||||||:::|||||||||||||:::
Db  122 LeuGluAlaLeuArgArgLeuLysValGluGlyHisArgPheProArgThrIleHisMet 141

Qy  556 TCTCTGGGCCATGATGAGGAGGTCATCAGGGACAGGGGCTCAGAGGATCTCAGCCCTGCTA 615
            |||||||||||||||||:::|||
Db  142 ThrPheValProAspGluGluValGlyGly 151

Qy  616 CAGTCAAGGGCGGTCCAACTA------GCCTTCAGTGTGGACGAGGGGGGC 660
            :::||:||:||||||||||:||||||
Db  152 --HisGlnGlyMetGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170

Qy  661 TTCATCTTGGATGATTTCATTCCTAACTTCAAGAAGCCCATGCCCTTGATTGCAGTC--- 717
            |||||||||:::|||||| |||
Db  171 PheAlaLeuAspGluGluGlyIleAlaAsn------ 187

Qy  718 ---TCAGAGAGGGTTCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCACTCT 774
            |||||||:::||:||:||||||||||||
Db  188 TyrSerGluArgSerProTrpTrpProValArgValThrSerThrGlyArgProGlyHisHisAla 207

Qy  775 TCAGCTCCTCCAAAGGAGGACAAGAAGCATTGGCATCCTTGCTGCTGCTCAGCCGATTGGAG 834
            :::||||||||||:::||||
Db  208 SerArgPheMetGluAspThr------ 220
```

```
Db  127 ProGluPheHisAlaLeuArgAlaGlyPheAlaLeuAspGluGlyIleAlaAsn------ 144
            |||||
Qy  694 AAGCCACCATCGCCTTGATTGCAGTC------TCAGACAAGGGTTCCATGAACCTCATGCTG 747

Db  145 ---ProThrAspAlaPheThrValPheTyrSerGluArgSerProTrpTrpProValArgVal 163
            |||   :::
Qy  748 CAAGTAAACATGACTTCAGGCCACTCTTCAGCTCCTCCAAAGGAGGACAAGAAGCATTGGCATC 807

Db  164 ThrSerThrGlyArgProGlyHisHisAlaSerArgPheMetGluAsp:Thr------ 179
            |||||||:::::|||
Qy  808 CTTGCACCTGCTGCTCAGCCGATTGGAGCAGAGACCAACACCCTATCATATATTGGAAGCGGG 867

Db  180 ------AlaAlaGluLysLeuHisLys------ 186
            |||
Qy  868 ACAGTGGTGACTGTATTGCAGCAAATGAGTTTCCCTTCCCTGTCAATATAATC 927

Db  187 ValValAsnSerIleLeuAlaPheArgGluLysGluTrp------ 202
            |||||:::|||
Qy  928 CTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAATCCCTTA 987

Db  203 GlnSerAsnProHisLeuLysGluGlySerVal------ 213
            ||||||||||||:::
Qy  988 ACCAATGCAATAATCAGGACCACCACGGCACTCACCATATTCAAAGCAGAGGGGTCAAGTTC 1047

Db  214 ------ThrSerValAsnLeuThrLysGluGlyIleGlyValAlaTyr 227
            :::
Qy 1048 AATGTCATCCCCCAGTGGCCCCAGCCACCCTCTGATGACAGGACA 1107

Db  228 AsnValIleProAlaThrMetSerAlaSerPheAspPheAspValAlaProValAlaAsp 247
            ||||||:::|||||:::
Qy 1108 GTCCAAGAGGTCCTAGAACTCACGAGAGAAC------AATTGTGGCTGATTAACAGAGTCCAG 1161

Db  248 PheLysAlaPheGluGluGlnLeuGlnSerGlnGlnAlaAlaGluGlyGluGlyValThr 267
            :::|||
Qy 1162 TTCCATGTGTTGAGTGCCTTTGACCCCCTCCCCGTTCAGCCCTTCTGATGACAAGGCCTTG 1221

Db  268 LeuGluPheHisLeuGlnLysTrpMetHisProGlnValThrProThrAspAspSerAsnPro 287
            |||:::|||||||||:::
Qy 1222 GGCTACCAGCTCTCCCCAGACCGTACAGTCCGTCTTCCCGGAAGTCAATATTACT--- 1278

Db  288 TrpTrpAlaAlaAlaPheSerArg------ValCysLysAspMetAsnLeuThrLeu 303
            :::|||||||||:::|||
Qy 1279 GCCCCAGTTACTTCTATTGGCAACACAGCAGCCGATTCTTTACAAACCTCACCACTGGC 1338

Db  304 GluProGluIleMetProAlaIleAlaThrAspAsnArgTyrIleArgAlaValGlyValPro 323
            |||||:::|||||
Qy 1339 ARCTACAGGTTCTACCCCATCTACTACACAGCCT------GAAGACTTCAAACGC 1386

Db  324 AlaLeuGlyPheSerProMetAsnArgThrProValLeuLeuHisAsnHisAspGluArg 343
            :::|||:::|||
Qy 1387 ATCCAT 1392

Db  344 LeuHisHis 345
            :::|||
Qy 1392
```

RESULT 4
US-08-555-860-4
Sequence 4, Application US/08555860
Patent No. 5585474
GENERAL INFORMATION:
APPLICANT: IWAKI, Kanso
APPLICANT: OHTA, Tsunetaka
APPLICANT: KURIOTO, Masashi
TITLE OF INVENTION: PROTEIN, DNA CODING SAID PROTEIN, AND
TITLE OF INVENTION: PREPARATION OF SAID PROTEIN
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: BRODY AND NEIMARK
STREET: 419 Seventh Street, N.W., Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20004
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/555,860
FILING DATE: 13-NOV-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/127,278
FILING DATE: 27-SEP-1993
APPLICATION NUMBER: JP 281136/1992
FILING DATE: 28-SEP-1992
ATTORNEY/AGENT INFORMATION:

NAME: NEIMARK, Sheridan
REGISTRATION NUMBER: 20,520
REFERENCE/DOCKET NUMBER: IWAKI-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-628-5197
TELEFAX: 202-737-3528
TELEX: 248633
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 408 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-555-860-4

Alignment Scores:
Pred. No.:
Score: 5.27e-10   Length: 408
Percent Similarity: 185.50   Matches: 89
Best Local Similarity: 39.14%   Conservative: 57
Query Match: 23.86%   Mismatches: 158
                       6.72%   Indels: 69
                       1   Gaps: 13
DB:

US-10-014-896-1 (1-1509) x US-08-555-860-4 (1-408)

Qy 319 CTGTTCACTATCCATCCAAGGCTCCGGACCCCCAGCTTGCAGCCTTACCTGCTGATGGCTCACTTT 378
            ||| ::::::|||
Db   62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLeuLeuAsnSerHisThr 81

Qy 379 GATGTGGGTGCCTGCCCCTGAAGAAGGCTCGGAGGGTGCCCCCCATTCTTCTGGGTTG---GAG 435
            :::||||||| |||
Db   82 AspValValProValPheLysGluHisTrpSerHisAspProPheGluAlaPheLysAsp 101

Qy 436 CGTGATGGGCGTCATCTTATGGTGCGGGGCCACACTGGACGACGAAGAGACTCTGTGATGGCATTA 495
            ::::||| ||||| |||
Db 102 SerGluGlyTyrIleTyrAlaAlaArgGlyAlaGlnAsnMetLysCysValSerIleGlnTyr 121

Qy 496 CTGCAGGCCTTGGAGCGTCCCTGCTGATCAGGAGAAGTACATCCCCGAAGAGATCTTTCTCATT 555
            |||||| ||| :::
Db 122 LeuGluAlaLeuArgArgLeuLysValGlyLysHisArgPheProArgThrIleHisMet 141

Qy 556 TCTCCGGCCCATGATGAGGAGTCATCAGGGACAGGGGCGTCAGAGGATCTCAGCCCTGCTA 615
            ||||||||| |||
Db 142 ThrPheValProAspGluValValGlyGly 151

Qy 616 CAGTCAAGGGCGGTCCCAGCTA-----------------------GCCTTCATTGTGACGAGGGGGC 660
            :::||||||||| |||
Db 152 ---HisGlnGlyMetGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170

Qy 661 TTCATCTTGATGATTCATTCCTAACTTCAAGAAGCCCATCGCCTTGATTGCAGTC--- 717
            ||| ||||||::: |||
Db 171 PheAlaLeuAspGluGlyIleIleAlaAsn------ProThrAspAlaPheThrValPhe 187

Qy 718 ---TCAGAGAAGGGTTCCATGAAACCTCATCGTCGCAAGTAAACATGAACGACTTCAGGCCACTCT 774
            ||||||||||:::|||
Db 188 TyrSerGluArgSerProTrpTrpValArgValThrSerThrHrGlyArgProGlyHisAla 207

Qy 775 TCAGCTCCTCCAAAGGACACAAGCAATGGCCATCCTTGCAGCTGCTGTCAGCCGGANTTGGAG 834
            ||| :::::::|||
Db 208 SerArgPheMetGluAspThr------------------------AlaAlaGlyLysLeuHis 220

Qy 835 CAGACACCAATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTG 894
            ||| :::::::|||
Db 221 Lys------------------------ValValAsnSerIleLeuLeuAlaPheArg 230

Qy 895 GCAAATGAGTTTCCCTTCCCTGTTCAATAATAATCCTGAGCAACCCATGGCTATTTGAACCA 954
            ||||||| :::::::|||
Db 231 GluLysGluTrp----------GlnArgLeuGlnSerAsnProHisLeuLysGluGly 246

Qy 955 CTTATAAGCAGGTTTATGGAGAGAAATCCCTTAACCAATGCAATAATCAGGACCACCACG 1014
            :::|||
Db 247 SerVal------------------------------ThrSerVal 251

Qy 1015 GCACTCACCATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCC 1074
            |||||| :::||||||||||||
Db 252 AsnLeuThrLysLeuGluGlyGlyValAlaTyrAsnValIleProAlaThrMetSerAla 271

Qy 1075 ACAGTCAACTTCCGGATTCACCCTGGACAGAGACGAGTCCTCAGAACTCACGAAG 1134
            :::||||||||| |||
Db 272 SerPheAspPheArgValAlaProAspPheLysAlaPheGluGlnLeuGln 291

Qy 1135 AAC------ATTGTGGCTGAACAGAGTCCAAGTTCCATGTGTGTTGAGTGCCTTGACCCC 1188
            :::
Db 292 SerTrpCysGlnAlaAlaGlyGluValThrLeuGluPheAlaGlnLnLysTrpMetHis 311

Qy 1189 CTCCCCGGTCAGCCCTTCTGATGACAAGGCCCTTGGGCTTACCAGCTGCTCCGCCAGACCGTA 1248
            :::||||||||||||:::||||||||||||
Db 312 ProGlnValThrProThrAspAspSerAsnProTrpTrpAlaLeuAlaPheSerArg 329

Qy 1249 CAGTCCGTGCTTCCCCGGAAGTCAATATTACT---GCCCCAGTTACTTCTATTGGCCAACACA 1305
            |||||||:::|||
Db 330 ----ValCysGlyAspMetAsnLeuThrLeuGluProGluIleMetProAlaAlaThr 347

Qy 1306 GACACGCCGGATTGTTTACAAAACCTCACCACTGGCATCTACAGGTTCTACCCCATCTACATA 1365
            ||| |||||||::::::|||
Db 348 AspAsnArgThrValTyrLeuArgAlaValGlyValProAlaLeuGlyPheSerProMetAsnArg 367

Qy 1366 CAGCCT------------------GAAGACTTCAAACGCCATCCAT 1392
            :::||||||:::|||
Db 368 ThrProValLeuLeuHisAspGluArgLeuHis 380

Left column (RESULT 5):

```
Qy  1015 GCACTCACCATATTCAAACAGAGGGTTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCC 1074
             |||||   :::  ||||||||| |||||||||||
Db   252 AsnLeuThrLysLeuGlyGlyValAlaTyrAsnValIleProAlaThrMetSerAla 271

Qy  1075 ACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGAAG 1134
             :::||||||:|| :||||:|::  ||:
Db   272 SerPheAspPheArgValAlaProAspPheLysAlaGluGlnLeuGln 291

Qy  1135 AAC------------ATTGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGACCCC 1188
              |||
Db   292 SerTrpCysGlnAlaAlaGlyGlyValThrLeuGlnLysTrpMetHis 311

Qy  1189 CTCCCCGTCAGCCCTTCGATGACAAGGCCTTGGGGCTTACCAGTCGTCCGCCAGACCGTA 1248
             ||||||:||||||:||||||||||
Db   312 ProGlnValThrProThrAspAspSerAsnProThrTrpAlaAlaPheSerArg------ 329

Qy  1249 CAGTCCGTCTCTTCCCGGAAGTCAATATTACT---GCCCCAGTTACTTCTTATTGGCAACACA 1305
             :||:||||:|| :||||:|:|| ||||||||||
Db   330 -----ValCysLysAspMetAsnLeuThrLeuGluProGluIleMetProAlaAlaThr 347

Qy  1306 GACAGCCGGATTCTTTACAAAACCTCACCACTGGCATCTACAGGTTCTACCCCATCTACATA 1365
             |||:||||||||||||||||||||
Db   348 AspAsnArgThrTyrIleArgAlaAlaValGlyValProAlaLeuGlyPheSerProMetAsnArg 367

Qy  1366 CAGCCT---------------GAAGACTTCAAACCGCATCCAT 1392
             |||                  :||::||||:|:|||
Db   368 ThrProValLeuLeuHisAspHisAspGluArgLeuHis 380
```

RESULT 5
US-09-814-951A-4
; Sequence 4, Application US/0981495lA
; Patent No. 6387661
; GENERAL INFORMATION:
; APPLICANT: SHAO, Wei et al
; TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
; ACID MOLECULES ENCODING HUMAN AMINOACYLASE, AND USES THEREOF
; FILE REFERENCE: CL001179
; CURRENT APPLICATION NUMBER: US/09/814,951A
; CURRENT FILING DATE: 2001-03-23
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 4
;   LENGTH: 408
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-814-951A-4

Alignment Scores:
Pred. No.:
Score: 5.27e-10   Length: 408
                    185.50   Matches: 89
Percent Similarity: 39.14%   Conservative: 57
                    23.86%   Mismatches: 158
Best Local Similarity: 6.72%   Indels: 69
Query Match: 4   Gaps: 13
DB:

US-10-014-896-1 (1-1509) x US-09-814-951A-4 (1-408)

```
Qy   319 CTGTTCACTATCCAAGGCTCGGACCCTGCTGCTTGATGGCTCACTTT 378
             ::: ||                 |||||| :::||:||
Db    62 ValLeuThrTrpProGlyThrAsnProThrLeuSerSerIleLeuLeuAsnSerHisThr 81

Qy   379 GAGTGTGGTCGCCTCGCCCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTCTGGGTTG---GAG 435
             ::||||||:||||| ||| |||  :||||||
Db    82 AspValProValPheLysGluHisTrpSerHisAspProheGluAlaPheLysAsp 101

Qy   436 CCGTGATGCGGCGTCATCTATGGTCGGGGGCACAGCAACAGAACTTGTGAGGCAATTA 495
             ::||||||||| || |||  ||||||:|
Db   102 SerGluGlyTyrIleTyrAlaArgGlyAlaGlnAspMetLysCysValSerIleGlnTyr 121

Qy   496 CTTCGAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACAATCCCCGGAAGATCTTTCTCATT 555
             |||:|||:||| |:|| :::
Db   122 LeuGluAlaValArgArgLeuLysValGlyHisArgPheProArgThrIleHisMet 141
```

Right column (RESULT 6):

```
Qy   556 TCTCTGGGCCATGATGAGGAGTCATCAGGAGACAGGGGGCTCAGAGGGATCTCAGCCCTGCTA 615
             |||||||||||
Db   142 ThrPheValProAspGluGluValGlyGly------------------------------ 151

Qy   616 CAGTCAAGGGCGTGCCTCCAGCTA------------------GCCTTCATTGTGGAGGGGGC 660
             :::||||||| |||                              |||
Db   152 ---HisGlnGlyMetGluLeuPheValGlnArgProGluPheHisAlaLeuArgAlaGly 170

Qy   661 TTCATCTTGGATGATTTCATTCCTAACTTCAAAGAAGACCCCATCGCCTTGATTGCTGTC--- 717
             :::||||||||           |||
Db   171 PheAlaLeuAspGluLysIleLeuAlaAsn------------ProThrAspAlaPheThrValPhe 187

Qy   718 ---TCAGAAGAGGGTTCCATGAACCTCATGCTGCAAGTAAACAATGACTTCAGGCGCCACTCT 774
                :::||||||||||
Db   188 TyrSerGluArgSerProThrTrpValArgValThrSerThrGlyArgProGlyHisAla 207

Qy   775 TCAGCTCCTCCAAAGGACAAGCATTGGCATCCTTGCAGCTGCTCTGCGATTGGAG 834
             ||| ::: |||||||||
Db   208 SerArgProMetGluAspThr-------------------AlaAlaGluLysLeuHis 220

Qy   835 CAGACACCAAGCCTATCATCATATTTGGAAGGGGAGACAGTGGTGCAAGTGATTGCAGCAACTG 894
             |||||||||
Db   221 Lys----------------------ValAlaAsnSerIleLeuAlaPheArg 230

Qy   895 GCAAATGAGTTTCCCTTCCCGTCAATATAATCCTGAGCAACCCCATGGCGCTATTTGAACCA 954
             :::  ||||||||:|||||
Db   231 GluLysGluTrp------------GlnArgLeuGlnSerAsnProHisLeuLysGluGly 246

Qy   955 CTTATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCAAATGCAATAATCAGGACCACACG 1014
             |||:|
Db   247 SerVal------------------------------------ThrSerVal 251
```

RESULT 6
US-09-134-001C-4338
; Sequence 4338, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08

```
Qy  1015 GCACTCACCATATTCAAACAGAGGGTTCAAGTTCAATGTCATCCCCCAGTGGCCCAGGCC 1074
             |||||   :::  |||||||||::||||||||||
Db   252 AsnLeuThrLysLeuGlyGlyValAlaTyrAsnValIleProAlaThrMetSerAla 271

Qy  1075 ACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGAGGTCCTAGAACTCACGAAG 1134
             :::||||||:|| :||||:|::  ||:
Db   272 SerPheAspPheArgValAlaProAspPheLysAlaGluGlnLeuGln 291

Qy  1135 AAC------------ATTGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGACCCC 1188
              |||
Db   292 SerTrpCysGlnAlaAlaGlyGlyValThrLeuGlnLysTrpMetHis 311

Qy  1189 CTCCCCGTCAGCCCTTCGATGACAAGGCCTTGGGGCTTACCAGTCGTCCGCCAGACCGTA 1248
             ||||||:||||||:||||||||||
Db   312 ProGlnValThrProThrAspAspSerAsnProThrTrpAlaAlaPheSerArg------ 329

Qy  1249 CAGTCCGTCTCTTCCCGGAAGTCAATATTACT---GCCCCAGTTACTTCTTATTGGCAACACA 1305
             :||:||||:|| :||||:|:|| ||||||||||
Db   330 -----ValCysLysAspMetAsnLeuThrLeuGluProGluIleMetProAlaAlaThr 347

Qy  1306 GACAGCCGGATTCTTTACAAAACCTCACCACTGGCATCTACAGGTTCTACCCCATCTACATA 1365
             |||:||||||||||||||||||||
Db   348 AspAsnArgThrTyrIleArgAlaAlaValGlyValProAlaLeuGlyPheSerProMetAsnArg 367

Qy  1366 CAGCCT---------------GAAGACTTCAAACCGCATCCAT 1392
             |||                  :||::||||:|:|||
Db   368 ThrProValLeuLeuHisAspHisAspGluArgLeuHis 380
```

```
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4338
;   LENGTH: 431
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4338
```

Alignment Scores:
```
Pred. No.:
Score:                   3.13e-08
                         168.00
Percent Similarity:      34.85%
Best Local Similarity:   21.24%
Query Match:             6.08%
DB:                      4
```

```
Length:        431
Matches:       103
Conservative:  66
Mismatches:    196
Indels:        120
Gaps:          20
```

US-10-014-896-1 (1-1509) x US-09-134-001C-4338 (1-431)

```
Qy   121 TCTCAGTTCAGCAAAGAGGAACGCGTGCGATGAGAAAGAGGGCCCTGAAAGGTGCCATCCAG 180
             ||      ||   ||          |||           |||
Db    21 ServalLeuSerAsnGluGluArgVal------------GluIleLeuSerAspIleValSer 37

Qy   181 ATTCCACACAGTGACTTTTAGCTCTGAGAAGTCCAATACTACAGCCCTGGCTCAGTTCGGA 240
         ||||||      |||                          |||
Db    38 IleLysThrVal------------------------------AsnSerAsnGluLeu---GluValAla 49

Qy   241 AAATACATTCATAAAGTCTTTCCTACAGTGGTCAGACCAGCTTTATCCAGCATGAAGTC 300
         ||||||    |:  :|||||              |:::|||::|||
Db    50 GlnTyrPheGluArgLeuPheSerGlnTyrGlyIleArgSerTyrIleAspIleValAla 69

Qy   301 GTGGAAGAGTATAGCCACCTGTTCACTATCCAAGGCTGGACCCCAGCTTGCAGCCCTAC 360
          |||       |||                      |||          |||
Db    70 AspGlyArgGlyAlaAsnLeuIleAlaThrValGlySerSerHisProValIleGly 87

Qy   361 CTGCTGATGGGTCTCACTTTGATGTGGGTG---CCTGCCCCTGAAGAAGGGCTGGGAGTGCCC 417
            :||                   :||  :||          |||||||
Db    88 ---IleSerGlyHisMetAspValSerGluGlyAsnHisAspTrpThrThrTyrAsp 106

Qy   418 CCATTCTCTGGGTTGGAGCCGTCATCATTGGGGGCCCACACTGGACGACAAG 477
         ||||:||:|  :||              |||      |||
Db   107 ProPheThrLeuThrGluAspGlnGlyIleTyrLeuTyrGlyArgGlyValAlaAlaAspMetLys 126

Qy   478 AACTCTGTGATGGGCATTACTGCCAGGCCTCCTGCTGATCATCAGGAAGTACATCCCCC 537
                                       |||
Db   127 SerGlyLeuAlaAlaLeu-----AlaIleAlaLeuIleLeuIleLeuGluSerGlyLys 144

Qy   538 CGAAGA-------------------TCTTTCTTCATTTCTCTGGGCCATGAGTGAGGAGTCACAGGG 585
            :::                    |||||:|||   ::||:|       |||||
Db   145 LeuThrGlnGlyThrIleLysPheMetAlaThrValGlyGluMetGluGlnSerGly 164

Qy   586 ACAGGG--------------------GCTCAGAGGATCTCAGCCCTGCTCACAGTCAAGG 624
         :::                       ||:|||||||||||||||:::
Db   165 SerGlnIleLeuPheGluLysGlyTyrAlaAspAspLeuAlaLeuLeuIleAlaGlu 184

Qy   625 GGCGCGTCCAGCTAGCCTTCATTGTGGACGAGGGGGCTTCATTCGATGATTTCATTCCT 684
                                    |||
Db   185 ------------------------------------------Pro 185

Qy   685 AACTTCAAGAGCCCATCGCCTTGATTCGAGTCTCCATGAGAAGGGTTCCATGAACCTCATG 744
         :: |||          |||||||||||||||:::
Db   186 SerPheProSer--------------LeuValTyrAlaHisLysGlySerMetAspPheArg 201

Qy   745 CTGCAAGTAAACATGACTTCAGGCCACTCTTCAGCTCCTCCAAAAGGAGACAAGCATTGGC 804
         ||||||    |||          :::                 |||
Db   202 IleLysSerLysGlyArgAlaSerHisSerSerIlePro------------------ 214

Qy   805 ATCCTTGCCAGCTGCTTGTCAGCCGATTGGAGGACGACACCAATGCCTATCATATTTGGAAGC 864
            ::             ||                           |||
Db   215 ------------------------------------------------PheLeuGlyGln 218

Qy   865 GGGACAGTGGTTGACTGTATTGCAGGCAAACTGGCCAAAT-------------GAGTTTCCTTCCCT 915
         :::              ||| :|::             :||::|
Db                                                    -----PheLeuGlyGln 218
```

```
RESULT 7
US-09-134-001C-5125
; Sequence 5125, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
;   APPLICANT: Lynn Doucette-Stamm et al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
;   TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: GTC-007
;   CURRENT APPLICATION NUMBER: US/09/134,001C
;   CURRENT FILING DATE: 1998-08-13
;   PRIOR APPLICATION NUMBER: US 60/064,964
;   PRIOR FILING DATE: 1997-11-08
;   PRIOR APPLICATION NUMBER: US 60/055,779
;   PRIOR FILING DATE: 1997-08-14
;   NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 5125
;   LENGTH: 446
;   TYPE: PRT
;   ORGANISM: Staphylococcus epidermidis
US-09-134-001C-5125
```

Alignment Scores:
```
Pred. No.:
Score:                   7.27e-07
                         154.50
Percent Similarity:      36.03%
Best Local Similarity:   23.33%
```

```
Length:        446
Matches:       101
Conservative:  55
Mismatches:    160
```

```
Qy   219 AsnAlaIleLysProLeuLeuGluProHeIleGlnAsnIleAsnGlnGluTyrGluLysIle 238
         |||
Db   219 GlnLeuProAsnHisIleThrLysGluGluGluLysSerValAlaArgThrIleGlyIleValMet 278

Qy   239 MetGlnThrValLysGlyGluSerLeuAspPheSerAsnMetIleAsnLysLeuGluAsn 258
         :::|||||:|:::
Db   259 AlaThrAlaGluProAsnValArgThrIlePheGluGluThrIleAsnAsnLysValLysAsnIle 315

Qy   259 GlnLeuProAsnHisIleThrLysGluGluGluLysSerValAlaArgThrIleGlyIleValMet 278
         :::|||
Db   279 ThrAsn-------SerLeuValGlnGlyValThrGlnGlnValThrGlnGlnValGlyThrValAsn 295

Qy   279 ThrAsn-------SerLeuValGlnGlyValThrGlnGlnValThrGlnGlnValGlyThrValAsn 295
            :::|||           |||
Db   296 AlaThrAlaGluProAsnValArgThrIlePheGluGluThrIleAsnAsnLysValLysAsnIle 315

Qy   296 GCCAGGCCACAGTCAACTTCCGGATTCACCCTGGACAGACAGTCCAAGGAGGTCCTGAGTT 1287
         :::
Db   316 LeuPheAsnLysTyrValGlyIleGlnAlaAsnHisAsnGlyAlaSerLeuThrGlnGluLeu 335

Qy   316 LeuPheAsnLysTyrValGlyIleGlnAlaAsnHisAsnGlyAlaSerLeuThrGlnGluLeu 335
         |||
Db   336 TyrLeuAspLeuGluProValValThrThrGlyGlnAsnArgLeuValGlyLeuGlyPhe 355

Qy  1177 GCCTTGACCCCCTCCCGGTCAGCCCGTTCTGATGAAGG-----GCCTTGGGCTAC 1227
         :::                                        |||||:::
Db  1228 CAGCTGCTCCGCGCAGACCGTACAGTCCGTCTTCCGGAAGTCAATATTACTGCCCAGTT 1287

Qy  1228 CAGCTGCTCCGCGCAGACCGTACAGTCCGTCTTCCGGAAGTCAATATTACTGCCCAGTT 1287
         |||                  |:             |:||
Db  1288 ACTTCCTATTGGCAACACAGACGCGGATTCTTTACAAACCTCACCACTGCCATCTACAGG 1347

Qy  1288 ACTTCCTATTGGCAACACAGACGCGGATTCTTTACAAACCTCACCACTGCCATCTACAGG 1347
         |||          :::                    |||
Db  1348 TTCTACCCCAACTACATACAGACGCCTGAAAGACTTCAAAACGCATC---------- 1389

Qy  1348 TTCTACCCCAACTACATACAGACGCCTGAAAGACTTCAAAACGCATC---------- 1389
           |||:::|||          |||:::|||
Db  1390 -----CATGGAGTCAAACAGAGAAAATCTCAGTCCAAGCCTATGAGAGCCAAGTGAAATTC 1443

Qy  1390 -----CATGGAGTCAAACAGAGAAAATCTCAGTCCAAGCCTATGAGAGCCAAGTGAAATTC 1443
              :::|||          |||:::
Db  1444 ATCTTTGAGTTGATT 1458

Qy  1444 ATCTTTGAGTTGATT 1458
         |||
Db  1390 GlyProHisGlnIleAsnGlyCysValGlyLysValAlaAsnTyrLeuGluValGluTyr 419

Qy   400 GlyProHisGlnIleAsnGlyCysValGlyLysValAlaAsnTyrLeuGluValGluTyr 419
         |||
Db   420 TyrIleGluPheIle 424
```

```
Query Match:          5.59%      Indels:     117
DB:                   4          Gaps:       20
```

```
Qy  US-10-014-896-1 (1-1509) x US-09-134-001C-5125 (1-446)

Qy   364 CTGATGGCTTCACTTTGATGTGTTGGTCGCCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCATTC 423
         ::::  ||||||| ::|||||::|||||||||||::|||||||||| ::::|||||| |||||
Db    58 IleLeuCysHisValAspValValProAla---GlyAspGlyThrProSerAsnProPhe 76

Qy   424 TCTGGGTTGGAGCGTGATGGCGTCATCTATGGTCGGGGCCACTGGACAAGAAGAACTCT 483
         ::       ||| ||||||||||::||||||||||||||||||||
Db    77 GlnProValValThrArgAspAlaIleIleLeuAlaArgGlyThrLeuAspAspLysGlyPro 96

Qy   484 GTGATGGCCAATACTGCAGGGCCTTGGAGCTCCTGCTGATCAGGAGGTACAACCCCGAACA 543
         ::||| |||  :::||||||:::|||| |||||||::::
Db    97 ThrIleAlaAlaThrTyrTyrAlaValLysIleLeuAsnGluMetLysValAspTrpLysLys 116

Qy   544 TCTTTCTTCATTTCTTCTCTGGGCCATGATGAGGAGGAGTCATCAGGAGGGGCCTCAGAGG--- 600
         |||    ||||||||||||||::||||||||||||||||||||
Db   117 ArgIleHisIleIleIleGlyThrAspGluGluSerProLysCysThrAspArgTyr 136

Qy   601 --------------------------------ATCTCAGCCCTG 612
                                         :||||||::::
Db   137 PheLysThrGluGluMetProThrLeuGlyPheAlaProAlaGluProAlaIle 156

Qy   613 CTACAGTCAAGGGGCGTCCAGCTACCCTTCATTGTG---------------------- 648
         ::|||::||||::|:|||||| ||||||:::|| |||
Db   157 HisGlyGluLysGlyIleThrThrPheAspLeuValGlnAsnGluValThrGluAspThr 176

Qy   649 GACGAGGGGGGGCTTCATCTTGGATGATTTC----------------------ATTCCT 684
         ||| ||||||||||||:::::||||||||| ::
Db   177 AspGluProAspTyrGluLeuLeuLysPheGluSerGlyGlnAsnArgTyrAsnMetValPro 196

Qy   685 AACTTCAAGAAGCCCATCGCCTTGAATTGCAGTCTCAGAGAAG------GGT 729
         :::::||| ::: ||| |||||| ||| ||| :::
Db   197 AspTyrAlaLys------AlaGluValLeuValLysGluAsnMetThrAspValIleGln 214

Qy   730 TCCATCAAACCTCATCGTGCAAGTAAACATCAGACTTCAGGCACCACTCTTCAGCTCCTCCAAAG 789
         :::||||| :::|| |||||::|||| |||||||||::|||||||||||
Db   215 AsnPheGluAsnPheLeuGlnGlnAsnGlnLeuGlnGlyGluSerThr----------- 230

Qy   790 GAGACAAGCAATTGCATCCTTGCACCTGCTGTCCGATTGGGAGCAGCAGCAGCAATGGCCT 849
         |||||||||||||||||||:::||||||||||||||||||||||||||| |||
Db   231 ---ValAspSerGlyIleLeuIleLeuThrIleGluGlyLysAlaValHisGlyMetAsp 249

Qy   850 ATCATATTTGGAAGCGGGGACCAGTGTGACTGTATTGCCAGCAACTGGCAAAT-------- 900
         :::    :::||||||::::||||| |||||||||||:::
Db   250 ProSerLeuGlyValAsnAlaGlyLeuPheLeuLeuLysPheLeuAlaSerLeuAsnLeu 269

Qy   901 --------------CAGTTTCCCTTCCCGTCAATATAATCCTGAGCAACCATGGCTA 945
               ::: |||||||::|||||||||| :::|| |||
Db   270 AsnLysSerAlaLysAspPhe-----------------ValGluPheAsnGluArgTyrLeu 284

Qy   946 TTTGAACCACTTATAAGCACAGTTTATGGAGAGAAAATCCCCTTA----------ACAAATGCA 996
         |||||||||||:::|||||::::|||||||||||||| :::
Db   285 PheGlu------SerHisPheGlyLeuGluLysMetGlyMetLysPheHisThrAspIle 301

Qy   997 ATAATCAGGACCACCACGGGACTCACCATA--------------- 1026
         :::||||||||||||||||||||||
Db   302 MetGlyAspValThrThrAsnIleGlyValIleSerTyrAspLysLysGlyLysAlaGlySer 321

Qy  1027 -------------TTCAAAGCAGGGGTCAAGTTCAATGTCATCCCCCAGTGGCC 1068
                     |||||||::::||||| :::|||||||||| |||
Db   322 TyrGlyIleAsnLeuArgTyrProGluGlyPheLysPheGlu----------- 335

Qy  1069 CAGGCCACAGTCAAGTCCGGATTCACCCGGGACACAGAGGTCCTAGAACTC 1128
         ||| :::|||||::||||||||||||||||| :::||||||::||| |||
Db   336 AspAlaIleAlaAspArgPheArg--------SerGluIleAsnGluLeu 348

Qy  1129 ACGAAGAACATTGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGTGCCTTTGACCCC 1188
         |||||||||||:::||||::|||||||||||||::|||||||||||
Db   349 GlyPheAsnLeuLeuGluLeuGlyLysValGlnLysProHisTyrValAspLysLysAsnAspPro 368
```

```
RESULT 8
US-09-575-602-4
; Sequence 4, Application US/09575602
; Patent No. 6392123
; GENERAL INFORMATION:
;  APPLICANT: Crossland, Lyle D
;  APPLICANT: Harper, Stacy M
;  TITLE OF INVENTION: Method of Hybrid Seed Production Using
;  TITLE OF INVENTION: Conditional Female Sterility
;  NUMBER OF SEQUENCES: 14
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: No. 6392123artis Corporation - Patent & Trademark
;  ADDRESSEE: Dept.
;  STREET: P.O. Box 12257
;  CITY: Research Triangle Park
;  STATE: NCNY
;  COUNTRY: USA
;  ZIP: 22057
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: Patentin Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09,575,602
;  FILING DATE:
;  CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;  APPLICATION NUMBER: 09/032,086
;  FILING DATE:
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Pace, Gary M
;  REGISTRATION NUMBER: 40,403
;  REFERENCE/DOCKET NUMBER: CGC 1915/Reg
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 919-541-8582
;  TELEFAX: 919-541-8689
;  INFORMATION FOR SEQ ID NO: 4:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 383 amino acids
;  TYPE: amino acid
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-09-575-602-4
```

```
Qy  1189 CTCCCCGTCAGCCCTTCTGATGACAAGGCGTTGGGCTACCAGTGTGCTCCGGCCAGACCGTA 1248
          |||       ||| |||  |||
Db   369 PheVal----------------------LysThrLeuValAsnAlaTyrArgAspGlnThr---- 381

Qy  1249 CAGTCCGTCTTCCCGGAAGTCAAATATTACTGCCCCAGTTACTTCTATTGGCAACAGACAGAG 1308
          :::::|| ::: |||||| ||| |||||| ::||| ||||| |||
Db   382 --------------------GlyAspMetThrGluProTyrThrIleGlyThrTyr 395

Qy  1309 AGCCGGATTCTTTACAAACCTCACCACACTGGCATCTACAGGTTCACCACCATCTACATACAG 1368
          :::|||||||::|||||||| |||  :::||||||| :::|||||||
Db   396 AlaArg------AsnLeuAspLysGlyVal--AlaPheGlyAlaMetPheAlaAsp 411

Qy  1369 CCTGAAGACTTCAAACGCATCCATGGAGTCAACGAGAAAAATCTCA------GTC 1416
          ||||| :::||| :::|||||||::::||||| |||||| :::
Db   412 SerGluAspLeu----MetHisGlnLysAsnGluTyrIleThrLysLysGlnLeuIle 429

Qy  1417 CAAGCCTATCGAGACCCAAGTGAAATTCATCTTTGAGTTG 1455
          |||::: |||||| :::|||||::: |||||| |||
Db   430 AsnAlaThrSerIleTyrLeuGluAlaIleTyrArgAlaLeu 442
```

US-10-014-896-1 (1-1509) x US-09-575-602-4 (1-383)

```
Qy   361 CTGCTGAATGGCTCACTTTGATGTGGTTGCCTGCCCCTGAGAAGGCTGGGAGGTGCCCCCA 420
         |||||||    ||||||    |||     ||::: |||||
Db    76 LeuLeuLeuAlaGlyHisThrAspThrValProPheAspAspGlyArgTrpThrArgAspPro 95

Qy   421 TTCTCTGGGTTGGAGCCTGATGGCGTGGAGCGTCTATGGGTCGGGGCACACTGGACGAAGAAC 480
         |||||||    |||||||     ::||||||    |:|||||||    |||
Db    96 PheThrLeuThrGluHisAspGlyLysLeuTyrGlyLeuGlyThrAlaAspMetLysGly 115

Qy   481 TCTGTGATGGCATTACTGCAGGCCTTGGAGCTCCTCGATCAGGAAGTACATCCCCGA 540
         |||||| |||||  :: ||||||  ::: |||
Db   116 PhePheAlaPheIleLeuAspAlaLeuArgAspValThrLysGlyLeu 133

Qy   541 AGATCTTTCTTCATTTCTCTGGGCCATGATGAGGAGTGCACAGGGGCTCAGAGG 600
         ||||||   ||||||||||    :||||||    |||
Db   134 LysProLeuTyrIleLeuAlaThrAlaAspGluGluThrSerMetAlaGlyAlaArgTyr 153

Qy   601 ATCTCAGCCCTGCTACAGTCAAGGGCGTCCAGCTACGCTTCATTGTGACGAGGGGGGC 660
         |||:::   ||||||    |||
Db   154 PheAlaGluThrThrAlaLeuArg------ProAspCysAlaIleIleGlyGlu 169

Qy   661 TTCATCTTGGATGATTCATTCCTAACTTCAAGAAGCCCATCGGCCTTGAGTTGCAGTCTCA 720
         |||             ::: |||
Db       ----------------------ProThrSerLeuGlnProValArg------Lys 133

Qy   721 GAGAGAGGGTTCCATGAAACCCATGCTGCAAGTAAACATGAGGACTTCAGGCCACTCTTCAGCT 780
         ||||||||||    :::|||
Db   170 ------------------------ProThrSerLeuGlnProValArg----Ala 178

Qy   781 CCTCCAAAGGAGGACAAGACCATTGGCATCCTTGCAGCTGCGTGTCAGCCGGATGGAGCAGACA 840
         |||:::   |||    |||    |||
Db   179 HisLysGlyHisIleSerAsnAlaLeuIleAlaArgIleGluGlnGlySerGlyHisSerSerAsp 198

Qy   901 GAGTTT---------- 939
         :::
Db   226 ArgThrTyrHisTyrGluAlaPheThrValProTyrPro------ 237

Qy  1000 ATCAGGAGGACACCACGCACTCACCATATTCAAAGCAGGGGTCAAGTTCAATGTCATCCCC 1059
         |||
Db   237 ------------------GlyHisSerIleLeuGlnLeuArgAspAsnLeuLysGlu 225

Qy  1060 CCAGTGGCCCAGGCCACAGTCAACTTCACCCTGGACAGACAGTCCAAGAGGTC 1119
         |||:::    ||||||    |||||||    |||
Db   255 AlaCysCysGluLeuHisMetAspIleArgProLeuProGlyMetThrLeuAsnGluLeu 274

Qy  1120 CTAGGAACTCACGAGACAACAATTGTGGCTGATAA---------- 1151
         |||    :::||||||    |||
Db   275 AsnGlyLeuLeuAsnAspAlaLeuAlaProValSerGluArgTrpProGlyArgLeuThr 294

Qy  1152 ---------CAGAGTCCAGTTCCAATGTGTTGAGTGCCTTGACCCCCTCCCCGTCAG 1199
         :::   ||||||||| |||
Db   295 ValAspGluLeuHisProProIleProGlyTyrGluCys------ProProAsnHisGln 312

Qy  1200 -----CCCTTCTGATGACAAGGGCCTTGGGCTACCAGGCCTACACTGC 1253
         |||||||    :::
Db   313 LeuValGluValGlyAlaValGlyGlyGluGlyValAlaLysThrGluValGlyValAlaAsnTyrCysThr 332

Qy  1254 CGTCTTCCGGAAGCTCAATATTACTGCCCCAGTTACTTCTATTGG----------- 1298
         |||    ||||||||||:::
Db   333 GluAlaProPheIleGlnThrLeuCysProThrLeuValLeuGlyProGlySerIleAsn 352

Qy  1299 CAACACAGACAGCGGATTCTTTAC--------------AAACCT---------- 1328
         |||    :::||||||||    |||
Db   353 GlnAlaHisGlnProAspGluTyrLeuGluThrArgPheIleLysProThrArgGluLeu 372
```

Alignment Scores:
Pred. No.:
Score: 0.00277
                           118.50
Percent Similarity: 32.34%   Length: 389
Best Local Similarity: 18.09%   Matches: 85
Query Match: 4.29%   Conservative: 67
                                          Mismatches: 167
                                          Indels: 151
                                          Gaps: 19

DB:

US-10-014-896-1 (1-1509) x US-09-134-001C-3161 (1-389)

```
Qy   241 AAATACAATTCAT-----------AAAGTCTTTCCTACAGTGGTCAGCCACCAGCTTTATCCAG 291
         |||||||||:|:             ||| ||| |||   :::||||::|:
Db     4 LysTyrIleTyrThrSerLysLysGlyValProTyrMetIleAsn-------Gln 19

Qy   292 CATGAAAGTTGGTGGAAAGATGTATAGCCACCTGTTCACTATACCAA-----GGCTCGGAC 342
         :::||||||    |||:::   ||||||
Db    20 LysArgLeuLeuAspCysPheLeuGluLeuValGlnIleIleAspSerGluThrGlyHisGlu 39

Qy   343 CCCAGCTTGCCAGCCCTACCTG--------- 363
         :::|||||||||||||
Db    40 GluThrIleGlnProTyrLeuLysAspThrPheGluLysMetGlyLeuHisLeuHisValIleGlu 59

Qy   363 ------------------ 363
         |||
Db    60 AspGluAlaSerLysAsnAsnArgLeuGlyAlaAlaAsnAsnLeuLeuCysThrLeuLysSer 79

Qy   364 ---------------------CTGATGGCTCCACTTTGAT----GTGGTG 387
         :::|||    |||
Db    80 AsnIleSerHisGlnAsnValProLysIleTyrPheThrSerHisMetAspThrArgValVal 99

Qy   388 CCTGCCCCTGAAGAAGGCTGGGGAGGTGCCCCCATTCTCTGGGTTGGAGCGTGATGGCGTC 447
         ||||||||||||||||    :::||||||||||    |||
Db   100 Pro----------------------ValLysGluAspProVal 113

Qy   448 ATCTATGGTCGGGGCCACA-----CTGGACGACAGAGAACTCTGTGATGGCATTA 495
         |||:||    |||    |||||   |||
Db   114 ValTyrSerAspGlyThrThrIleLeuAlaAspAspLysAlaGlyLeuAlaAlaIle 133

Qy   496 CTGGAGGCCTTGGAGCTCCTGCTGATCAGGAAGTACATCCCCGAGAGATCTTTCTCATT 555
         ||:::||||||    ||||||    |||
Db   134 IleGluAlaIleLeuLysGlnIleLeu-----GluSerAsnLeuProHisGlyIleGlnIleIle 152

Qy   556 TCTCTGGGCCCATGATGATGAGGAGTCATCAGGAGACCAGGGCGCTAGGAATCTCAGCCCGCTA 615
         |||||||||||||    :::||||||||||    |||
Db   153 IleIleThrValGlyGluGluValGlySerGlyLeuValGlyAlaLysGlyLeuAlaAlaIleAspThrArgLeu 172
```

```
;  TITLE OF INVENTION: RECOMBINANT L-N-CARBAMOYLASE
;  FILE REFERENCE: RECOMBINANT L-N-CARBAMOYLASE, THEREWITH
;  CURRENT APPLICATION NUMBER: US/10,285,055
;  CURRENT FILING DATE: 1999-04-02
;  EARLIER APPLICATION NUMBER: DE 198 14 813.5
;  EARLIER FILING DATE: 1998-04-02
;  NUMBER OF SEQ ID NOS: 2
;  SOFTWARE: Patentin Ver. 2.0
;
;  SEQ ID NO 2
;    LENGTH: 412
;    TYPE: PRT
;    ORGANISM: Arthrobacter aurescens
;  US-09-285-055-2
```

```
;  US-10-014-896-1 (1-1509) x US-09-285-055-2 (1-412)

;  Alignment Scores:

Pred. No.:
Score:                      0.00456
Percent Similarity:         116.50          Length:          412
Best Local Similarity:      40.23%          Matches:         71
Query Match:                20.40%          Conservative:    69
DB:                         4.22%           Mismatches:      149
                            4                Indels:          59
                                             Gaps:            16
```

```
Qy    616 CAGTCAAGGGGCGGGTCCAGCTAGCCTTCATTGTGGACGAGGGGGCTTCATTCTTGGATGAT 675
              ::: ::    |||:|||   |||:|||  ::|||
Db    173 ------------LeuAspAlaAspPheGlyTyrAlaValAsp-- 182

Qy    676 TTCATTCCTAACTTCAAGAAGCCCATCGGCCTTGATTGCAGTCTCAGAGAAGGGTTCCATG 735
              |||     :::  |||:||  :||  ||| :::|||
Db    183 -------------AlaSerLysAspValGlyThrThrValIleGlyValAlaProThrGlnVal 198

Qy    736 AACCTCATGCTGCAAGTAAACATGACTTCAGGCCCACTCCTCTCAGCTCCCAAAGGAGACA 795
              |||      ||| |||  |||:|||
Db    199 LysIleTyrThrThrIleLysGlyLysThrAlaHisAlaSerThrProLysLysGlyIle 218

Qy    796 AGC---ATTGGGCATCCTTGCAGTCGTCTGCAGCCGGATTGGAGCACCAATGCCTATC 852
              :::  |||:|||||||:::|||||:||||
Db    219 SerAlaIleAsnIleAlaSerLysAlaIleSerArgMetLys---- 232

Qy    853 ATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCCTTC 912
              |||||||
Db    233 -----LeuGlyLnValAspAlaLeuThrThrAlaAsnIleGlyLysPheHis 248

Qy    913 CCTGTCAATATAATAACCTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATG 972
              :::||||  ||
Db    232 --------- 232

Qy    973 GAGAGAAATCCCTTAACCAATGCAATAACAGGACCACCACCAGTCCACCATATTCAAA 1032
              |||  :::
Db    233 --------LeuGlyInValAspAlaLeuThrThrAlaAsnIleGlyLysPheHis 248

Qy   1033 GCAGGGGTCAAGTTCATCCCCCCAGTGGCCCAGGCCACAGTCAACTTCCGGATT 1092
              |||:|||
Db    249 GlyGlySerAlaThrAsnIleIleAlaAspGluValThrLeuGluAlaAlaArgSer 268

Qy   1093 CACCCTGGACAGACAGTC---CAAGAGGTCTTAGAACTCACG 1131
              |||:|||:|:|||:|||
Db    269 HisAspAspGlnSerIleAsnLysGlnValLysHisMetLysGlyLThrProPheGluThrThr 288

Qy   1132 AAGAACATGTGGCTGATAACAGGACGTCCAGTTCCCATGGTGTTGAGTGCCTTTGACCCCTC 1191
              |||  :::||||||
Db    289 AlaAsnGluLeuGly--GlyGlnValGluValLeuValGlyLysSerTyrProGlyPhe 307

Qy   1192 CCGGTCAGCCCTTCGATGACAAGGCCTTGGGCTACCAGTGCTCCGGCCACGGTACAG 1251
              |||||||
Db    308 GluValSerGlnAla---AspLysValThrGlnTyrAlaIleSerSerGlyAlaLeuAlaLeu 326

Qy   1252 TCCCTCTTCCCGGAAGTCAATAATTACTGCCCCAGTGTACTTTCTATTGGCAACACAGACAGC 1311
              |||||||
Db    327 GlyLeuLysGlyAspThrCysIleAlaGly---GlyGlySerAspGly 341

Qy   1312 CGATTCTTTACAAACCTACACCTGGCCATCTACCCCATCAATACAGACCT 1371
              |||||||
Db    342 AsnIleMetAsnGlnTyr-----GlyIle---ProServalIleLeuGly 355

Qy   1372 GAAGACTTCAAACGGCCATCAACGAGAGAAAATCTCAGTCAAGCTCATGAGACC 1431
              |||||||
Db    356 ValGlyTyrGluAsnIleHisThrThrSerGluArgIleAlaIleAlaLysAspMetTyrMet 375

Qy   1432 CAAGTGAAATTCAICTTTGAGTTGATTCAG 1461
              :::|||:||
Db    376 LeuThrArgGlnIleIleLysIleIleGlu 385
```

```
RESULT 10
US-09-285-055-2
Sequence 2, Application US/09285055
Patent No. 6352848
GENERAL INFORMATION:
    APPLICANT: ALTENBUCHNER, JOSEF
    APPLICANT: MATTES, RALF
    APPLICANT: PIETZSCH, MARKUS
    APPLICANT: SYLDATK, CHRISTOPH
    APPLICANT: WIESE, ANJA
    APPLICANT: WILMS, BURKARD
    TITLE OF INVENTION: RECOMBINANT L-N-CARBAMOYLASE FROM ARTHROBACTER
        AURESCENS AND METHOD OF PRODUCING L-AMINO ACIDS
    TITLE OF INVENTION:
```

```
Qy    331 CAAGGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGGTGCCT 390
              ::||||||:||:||||||| ::|||   |||   |||||||:|||||||||  :::||||||
Db     68 GluGlyThrAspProGluLeuProAlaIleAlaValGlySerHisSerPheSerVal--- 86

Qy    391 GCCCCTGAGCAAGAGGCTGGGAGGTGCCCCATTCTCTGGGTTGGAGGTGGCGTCATC 450
              |||:|||:||  ||||||  |||  |||   :::|||:||  ||:|||
Db     87 ---------------------ArgAsnGly----- 89

Qy    451 TATGGTCGGGGCGCACACTGACGACAAGAACTCTGATGAGGCATTACTCGCAGGGCCTTGGAG 510
              :::|||  :::|||||||
Db     90 ----------GlyMetPheAspGlyThrAlaGlyValCysAlaAlaLeuGluAlaAlaArg 106

Qy    511 CTCCTGCTGATCAGGAAGAGTACATCCCCGAAGGTCTTC---TTCATTTCTCTGGGCCAT 567
              :::||||||  :::||||||||||
Db    107 ValMetLeuGluAsnGlyTyrValAsnArgHisProPheGluPheIleAlaIleLeuValGlu 126

Qy    568 GATGAGGAG------TCATCAGGG---ACAGGGGGCTCAGAGGAGGATCTCAGCCCCTGCTA 615
              |||||||:|||:|||
Db    127 GluGluGlyAlaAlaArgPheSerSerGlyMetLeuGlyValGlyArgAlaIleAlaGlyLeuVal 146

Qy    616 CAGTCAAGGGGGCGGGTCCAGCTAGCCTTCATTGTGGACGAGGGGGCTTCATTCTTGGATGAT 675
              |||||||
Db    147 AlaAspArg-----GluLeuAspSerLeuValAspGluValSerValArgGln 164

Qy    676 TTCATTCCTAACTTC---AAGAAGCCCATCGGCCTTGATTGCAGTCTCAGAGAAGGGTTCC 732
              |||||||
Db    165 AlaAlaThrAlaPheGlyLeuLysProGlyLeuGlnAlaLeuAlaLeuAlaArgSerAlaAla 184

Qy    733 ---ATGAACCTCATGCTGCAAGTAAACATGACTTCAGGCCCACTCCTCTCAGCTCCCAAAG 789
              |||::||||:|||
Db    185 AspLeuArgAlaPheIleGluHisIleGluGlnGly---ProIleLeuGluGlnGlu 203

Qy    790 GAGACAAGCAATTGGGCATCCTTGCAGTCGTCTGCAGCCGGATTGGAGCACGAACCAATGGCT 849
              |||||||
Db    204 GlnIleGluGlyValValThrSerIleValGly--ValArgGlnLeuArgValAlaAla 222

Qy    850 ATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAATGAGTTTCCC 909
              :::||||||
Db    223 ValLysGlyArgSerAspHisAlaGlyThrThrProMetHisGlnAspAlaLeu 242

Qy    910 TTCCCTGTCAATATAATAACCTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTT 969
              |||:::|||:::
Db    243 ValProAlaAlaLeuMetValArgGlu------ValAsnArgPhe 255

Qy    970 ATGGAGAGAAATCCCTTAACCAATGCAATAATAACAGGACCACCACGGTCAGTCACCATATTC 1029
              :::|||
Db    256 Val-----AsnGluIleLeuAlaAspGlyThrValAlaThrValAlaGlyHisLeuThrValAla 273

Qy   1030 AAAGCAGGGGGTCAAGTTCAATGTCATCCCCCCAGTGGCCCAGGCCACAGTCAACTTCCGG 1089
```

RESULT 11
US-09-134-001C-4302
Sequence 4302, Application US/09134001C
Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4302
; LENGTH: 430
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4302

Alignment Scores:
Pred. No.:
Score: 0.00745
Percent Similarity: 36.45%
Best Local Similarity: 22.41%
Query Match: 4.15%
DB:

                    Length: 430
                    Matches: 67
                    Conservative: 42
                    Mismatches: 129
                    Indels: 62
                    Gaps: 12

US-10-014-896-1 (1-1509) x US-09-134-001C-4302 (1-430)

Qy   325 ACTATCCAAGGCTCGGACCCCAGCTTGCAGCCTACCTGCTGATGGCTCACTTTGATGTG 384
         ::||| ||| ||| |||| :|| ||| ||||| ||| |||
Db   74 ThrLeuLysGlyHisAsnProLysAlaProLysLeuIleLeuAsnGlyHisValAspVal 93

Qy   385 GTGCCTGCCCCTGAAGAAGGC---TGGGAGGTGCCCCCATTCTCTGGGTTGGAGGGTGAT 441
         ||||: |||:::: |||||||||||:
Db   94 AlaSerValAspAspSerGlnThrTyrTrpGlnTyrTrpProProLysLeuThrAsnLysAsp 113

Qy   442 GGGCGTCATCTATGGTCGGGCGCACACTGGACAAGAGAACTCTGTGATGGCCATTACTGCAG 501
         ||||||||||| |||||||| ||||| |||
Db   114 GluTrpLeuTyrGlyArgGlyValSerAspMetLysGlyLysGlyMetSerSerLeuPheTyr 133

Qy   502 GCCTTGGAGCTCCTGCTGATCAGGAGGAGTACATCCCCGAAGATCTTTCTTCATTTCTCTG 561
         ||||| ||| ||| |||
Db   134 ValLeuGluGlnLeuLeuHisGlnAlaGlyHisArgProGluGlyAspIleIleIleValGlnSer 153

Qy   562 GGCCATGATGAGGAGTCATCAGGGACAGGGGGCTCAGAGGAATCTCAGACCCTTGCTACAGTCA 621
         ||||| ||| ||| ::|| :::
Db   (alignment continues)

RESULT 12
US-09-343-494-9
Sequence 9, Application US/09343494
Patent No. 6413741
; GENERAL INFORMATION:
; APPLICANT: Jegla, Timothy J.
; APPLICANT: Wickenden, Alan
; APPLICANT: ICAgen, Incorporated
; TITLE OF INVENTION: Human Elk, a Voltage-Gated Potassium Channel Subunit
; FILE REFERENCE: 018512-0013200US
; CURRENT APPLICATION NUMBER: US/09/343,494
; CURRENT FILING DATE: 1999-06-30
; EARLIER APPLICATION NUMBER: US 60/091,469
; EARLIER FILING DATE: 1998-07-01
; EARLIER APPLICATION NUMBER: US 60/116,621
; EARLIER FILING DATE: 1999-01-21
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1284
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; FEATURE:
; OTHER INFORMATION: Drosophila Elk (dElk; Eag (ether a go-go)-like K+
; OTHER INFORMATION: gene) protein
US-09-343-494-9

Alignment Scores:
Pred. No.:
Score: 0.021
Score: 113.00
Percent Similarity: 34.62%

                    Length: 1284
                    Matches: 90
                    Conservative: 63

```
Best Local Similarity:  20.36%                    Mismatches:  151
Query Match:             4.09%                     Indels:      138
DB:                                                Gaps:        18

US-10-014-896-1 (1-1509) x US-09-343-494-9 (1-1284)

Qy   384 GGTGCTGCCCCTGAAGAAGGCTGGGAGGTGCCCCATTCTTCTGGGTTGGAGCGTGATGG 443
             |||||    |||    ::    |||||||||||||||||||||
Db   773 GlyAlaSerProLeuHisAsnIleSerAsnSerProLeu------------------- 785

Qy   444 CGTCATCTATGGTCGGGGACACACTGGACGACAAGAACTCTGTGATGGCATTACTGCAGGC 503
                                                              |||
Db   786 ------------------------HisAlaThrArgSerProLeu------LeuGly 794

Qy   504 CTTGGAGCTCCTGTCGTCGATCAGGAAGTACATCCCCGAAGATCTTTCTTCATTTCTCTGGG 563
             :|||:|||:|| |||||||  |||       ||||||         |||||
Db   795 MetGlySerProArgAsnGlnArgLeuHis-GlnArgGlyArgSerLeuIleThrLeuAr 814

Qy   564 CCATGATGAGGAGACGAGAGTCATCAGGACACAGGGGCTCGAGGGATCTCAGCCCTGCTACAGTCAAG 623
              :||                                                     :|:
Db   814 gGluThrAsnLys------------------------------------------ArgHisAr 821

Qy   624 GGGCCGTCCAGCTAGCCTTCATTGTTGACGAGGGGGGCTTC--------------- 663
              :::                                        |||
Db   821 gThrLeuAsnAlaAlaCysSerLeuAspArgGlySerPheGluGluProGluProLeuGl 841

Qy   664 ----------------------------ATCTTGGATGATTTCATTCC 683
                                              |||:||||
Db   841 uGluGluGlnSerSerGlyGlyLysArgProSerLeuArgLeuAspSerGlnValSe 861

Qy   684 TAACTTCAAGAAGCCCATCGCCGCCTTGATTGCAGTCTCAGAGAAGGGTGTCCATGAACCTCAT 743
             ::|| :::|||||:|||||||::||||||||||||
Db   861 uGlyGlnLeuMetGlyGlyValGlyValLeuValAlaLeuAlaGluLeuAlaAlaMetGlnArgSerSe 881

Qy   744 GCTGCAAGTAAACATGACTTCA---------------GGCCACTCTTCAGCTCCT---CCAAA 788
             :::|||                 ||||||
Db   881 rThrLeuHisGlnAspValAlaGlnLeuSerAlaGlnValArgAsnAlaIleSerAlaIle 901

Qy   789 GGAGACAAAGCATT------------------------ 801
             |||||
Db   901 uGlnGluMetThrPheThrSerHisSerSerLeuLysPheProPr 921

Qy   802 ----------------GGCATCCTTGCA------ 833
                              :||:|||
Db   921 oAlaArgSerIleProAsnIleSerGlyValAlaGlyThrArgSerGlyLvaAlaValGl 941

Qy   834 GCAGACACCAATGCCTAICATATTTGGAAGCGGGACAGTGGTGCGACTGTATTGCAGCAACT 893
             :::||                    ||||:||||||||||||
Db   941 yGCATCGTGCAGCGGATTGGA------------------------------ 960

Qy   894 GGCAAAAGAGTTTCCCCTTCCCCGTCCAAT------------------------ 938
             :|||||||:                  |||:
Db   960 ATAAACCCTGACAACC 980

Qy   939 ATGGCTATTTGAA--------------------CCACTTATAAGCAGGGTTTATGGAGAGAAAAATCCCTT 986
             |||                          |||||||:::
Db   980 aSerSerLysAlaProSerProValGlyProLysLysThrMetThrSerArgSerGl 999

Qy   987 AACCAAATGCAATAATCAGGACACCACGGCACTCACCAATATTCAAAGCAGGGGTCAAGTT 1046
             ||||||:|||:||||                    |||:||
Db  1000 nThrAspPheTyrArgIleAspPheGluProThrPheGluValLeuAlaAsnPro-- 1014

Qy  1047 CAANGTCATNCCCCCCAGGTGGCCCCAGGTGCTCAACTTCCCGGATTCACCCGGACAGAC 1106
             |||                     |||                    ||:
Db  1014 ----ArgLeuValLeuGlyLeuLeuGlyIleGluProAlaIleLeuLysAs 1032

Qy  1107 AGTCCAAGAGAGGTCCTAGAACTCACGAAGAACACATTGTGGCTGATAACAGAGTCCAGTTCCA 1166
             :::                 |||               :||:
Db  1032 nGluMet-----AspLeuLeuGlnLnLysGlnThrLeuGlnIleSerProLeuAsnTh 1052

Qy  1167 TGTGTTGAGTGCC------------------------TTTGACCCCCTCCCCCGTCAGCCCTTC 1205
             |||                                 ||:     |||||
Db  1052 eThrSerSerAlaValProThrProGlyArgIleTyrProProLeuAspAspGluAsnSe 1072
```

```
Qy  1206 TGATGACAAGGCCTTGGGCGTACCAGCTGCTCCGCCAGACCGTACAGT---CCGTCTTCC 1262
             |::|||:|||                                    |||    ||||:|||
Db  1072 rAsnAspPheArg-TrpThrMetLysHisSerAlaSerHisHisSerCysLysSerT 1092

Qy  1263 GGAAGTCAATATTACTGCCCCAG------------------------TTACTTCTATTGGCAACAC 1304
             |::|||||:||                                 ||||||
Db  1092 hrAspAlaLeuLeuSerProGluGlnProProIleSerIleLeuProValAspAlaT 1112

Qy  1305 AGACACCGGATTCTTTTACAA------------------------------- 1324
             |||:|||:||||||
Db  1112 hrProAlaProSerValGlnGluValArgSerArgSerIleArgLysSerThrS 1132

Qy  1325 ------------------------ACCTCACCACTGGCATCTACAGGTTCT------ 1355
                                      ||||:||||||||||||:||
Db  1132 erGlySerAsnSerSerLeuSerSerSerSerSerSerAsnSerCysLeuValSerG 1152

Qy  1356 CATCTACATACAGCCTGAAGACTTCAAACGGCATCCATGGAGTCAACGAGAAAAATCTCAGT 1415
             |||||||:||||:||||:||:||||||||||||||||||||
Db  1152 lnSerThrGlyAsnLeuThrThrThrAsnAlaSerValHisCysSerSerAsnSerSerGlnS 1172

Qy  1416 CC 1417
             :::
Db  1172 er 1172
```

RESULT 13
US-08-479-722B-4
; Sequence 4, Application US/08479722B
; Patent No. 6074840
; GENERAL INFORMATION:
; APPLICANT: Bonadio, Jeffrey
; APPLICANT: Yin, Wushan
; TITLE OF INVENTION: LATENT TGF(BINDING PROTEIN (LTBP)
; TITLE OF INVENTION: GENES, COMPOSITIONS AND METHODS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Williams, Morgan & Amerson
;   STREET: 7676 Hillmont, Suite 250
;   CITY: Houston
;   STATE: Texas
;   COUNTRY: USA
;   ZIP: 77040
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/479,722B
;   FILING DATE: 07-JUN-1995
;   CLASSIFICATION:
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/316,650
;   FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US PCT/US95/02251
;   FILING DATE: 21-FEB-1995
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/199,780
;   FILING DATE: 18-FEB-1994
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Fussey, Shelley P.M.
;   REGISTRATION NUMBER: 39,458
;   REFERENCE/DOCKET NUMBER: 4100.000500/FUS
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (713) 934-7000
;   TELEFAX: (713) 934-7011
INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
;   LENGTH: 1253 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-479-722B-4

Alignment Scores:

Pred. No.:	0.0659
Score:	108.00
Percent Similarity:	32.09%
Best Local Similarity:	22.99%
Query Match:	4.02%
DB:	3

Length:	1253
Matches:	86
Conservative:	34
Mismatches:	143
Indels:	112
Gaps:	21

US-10-014-896-1 (1-1509) x US-08-479-722B-4 (1-1253)

```
Qy   1158 GACTCTGTTATCAGCCACAAATGTTCTTCGTGAGTCTCGAGACCTCTGGACTGTCTGTCC 1099
          ||||||||||::::||   ||||||||      ||||||     ::::::
Db    910 AspSerValLeuAlaThrAsnValThrGlnGln----------------------------- 920

Qy   1098 AGGGTGAATCCGGAAGTTGACTGGCCTGGGC---CACTGGGGGGGATGACATTGAACTT 1042
          ||| |||||  ||||||| |||||||   |||| :::|||||||||||||||||||||::
Db    921 -------------GluCysCysSerLeuGlyAlaGlyThrTrpGlyAspHisCysGluHis 936

Qy   1041 GACCCCTGCTTGAAATAGGGTGAGTGC---------------------------------- 1015
          |||||  ||||::|||                
Db    937 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg 956

Qy   1014 --CGTGGGTGGTCCTGATTATTGCATTGGTTGCTCAGGATTATATATTGACAGGAAGGAAAACTGCAT 958
             |||   ||||||||                ||||||||||||||||||||||||::
Db    957 LeuHisSerGlyGlnGlnHisCys---------------GluLeuCysIleProAlaHis 971

Qy    957 AAGTGGGTTCAAATAGCCCATGGGTTGCTCAGGATTATATTGACAGGAAGGAAAACTGCAT 898
          ||||||||||||| ||| :::|||||||||||||||||||| ::||||||||||||::
Db    972 ArgAspIleAspGluCysIleLeuPheGlyValAlaGluIleCysLysGluLysCysVal 991

Qy    897 TGCCAGTTGCTGCAATACAGTCACCACTGTCCCGCTTCC---AAATATGATAGGCATTGG 841
          |||   :::|||||                    ::||||||||||||||||||:
Db    992 AsnThrGlnProGlyTyrGluCysTyrCysLysGlnGlyPheTyrTyrAspGlyAsn--- 1010

Qy    840 TGTCTCGCTTCCAATCGGCCTGAC---AGCAGCTGCAAGGATGGCCAAATGCTTGTCT--- 791
          ||||||||||||     ::| ||| ::||||||||||||| |||:::|||
Db   1011 ---LeuLeuGluCysValAspGluCysLeuAspGluSerAsnCysArgAsnGly 1029

Qy    790 CCTTTGGAGGAGCTGAAGAGTGGCCTGAAGTGCATGTTTACTTGCAGCATGAGGTTCATGG 731
          ||| ||||||||||||||| |||||||  ::|||:::|||||||||||||||:
Db   1030 ValCysGluAsnThrArgGlyGlyTyrArg---CysAlaCysThrProProAlaGluTyr 1048

Qy    730 AACCCTTCTTCTGAGA---------CTGCAATCAAGGCGCATGGGCTGCTTCTTGAGTTAGGAAATGA 677
          :::|||:: |||      ||||||||||||| ||||||||||::|:::
Db   1049 SerProAlaGlnIleAlaGlnAsnCysLeuIleProLeuArgTrpSerThr------------ 1063

Qy    676 AATCATCCAAGATGAAGCCCCCTCGTTCCACAATGAAAGGCTAGCTGGACCCCCTTGACT 617
          ||| |||||||||||||||||||||||   ||||||||||||||||||||||||||||||
Db   1064 -----------------------------------------ProGln 1069

Qy    616 GTAGCAGGGCTGAGATCCTCTGAGCCCGTCCCTGATGACTCCTCATCATGGCCCAGAG 557
          ||||   ||| |||||                 ||| :::|||
Db   1070 CysAlaGlyValAlaSerGluGluArgThrArgGlyGlyTyrArg---CysAlaCysThrProProAlaGluTyr-- 1088

Qy    556 AAATGAAGAAGATCTTCGGGGAGTACTTCCGTATCACCAGAGACTCCAAGGCCTGCA 497
          |||||||   |||||||||||||| |||||||
Db   1089 -----AlaLeuThrPheAspAspCys---------------CysCysArgGlnProArgLeuGly 1103

Qy    496 GTAATGCCATCACAGAGTTCTTGTGCTCCAGTGTGCCCGACCATAGATGACGCCATCAC 437
          ::::|||||||||||                 |||||||||||||||||||||||::
Db   1104 -------------------------------ThrGlnCysArgPro---CysProProArgGlyThrGlySerGln 1117

Qy    436 GCTCCAAACCCAGAGAGAATGGGGGCA------------------------CCTCCCAGCCTTCTTCAGGGG 392
          ||| :::                                ||| |||
Db   1118 CysProThrSerGlnSerGluSerAsnSerPheTrpAspThrSerProLeuLeuGluGly 1137

Qy    391 CAGGGACCACATCAAAAGTGAG----------------------------------- 371
          :::::::|||                 ::
Db   1138 LysSerProArgAspAspSerSerGluGluGlyAspCysValSer 1157

Qy    370 ----------------CCATCAGCAGGTAGGGCTGCAAGCTGGGTCCGAGCCTGGATAGTGA 323
```

RESULT 14
; PCT-US95-02251-3
; Sequence 3, Application PC/TUS9502251
; GENERAL INFORMATION:
; APPLICANT:
;   TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR STIMULATING BONE
;   TITLE OF INVENTION: CELLS
;   NUMBER OF SEQUENCES: 18
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Arnold, White & Durkee
;   STREET: P.O. Box 4433
;   CITY: Houston
;   STATE: Texas
;   COUNTRY: United States of America
;   ZIP: 77210
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS/ASCII
;   SOFTWARE: Patentin Release #1.0, Version
;   SOFTWARE: #1.30
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: PCT/US95/02251
;   FILING DATE: CONCURRENTLY HEREWITH
;   CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: US 08/316,650
;   FILING DATE: 30-SEP-1994
;   CLASSIFICATION:
;   APPLICATION NUMBER: US 08/199,780
;   FILING DATE: 18-FEB-1994
;   CLASSIFICATION:
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Parker, David L.
;   REGISTRATION NUMBER: 32,165
;   REFERENCE/DOCKET NUMBER: UMIC009p--
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (512) 418-3000
;   TELEPHONE: (713) 789-2679
;   TELEFAX: (713) 789-0924
;   TELEX: 79-0924
;   INFORMATION FOR SEQ ID NO: 3:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 1251 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
; PCT-US95-02251-3

Alignment Scores:

Pred. No.:	0.0739
Score:	107.50
Percent Similarity:	29.97%
Best Local Similarity:	21.96%
Query Match:	4.00%
DB:	5

Length:	1251
Matches:	85
Conservative:	31
Mismatches:	133
Indels:	139
Gaps:	20

US-10-014-896-1 (1-1509) x PCT-US95-02251-3 (1-1251)

```
Db   1158 GlyArgCysValProArgProGlyGlyValAlaValCysGluCysPro--------------- 1172
          |||                       |||                           |||
Qy    322 ACAGGTGGGCTATACTCTTCCACGACACTTCATGCTGGATAAAGCTGGGTCTGGCTGACCACTGTAG 263

Db   1173 ---GlyGlyPheGlnLeuAspAlaSerArgAla--ArgCysValAspIleAspGluCysAr 1191
          ||| |||||||:: |||||::||||||||||||||:|||||::|||
Qy    262 GAAAGACTTTATGAAATGTTATTTTCCGAACTCAGCCAGGGGTGTAGTATTGGACTTCCAG 203

Db   1191 g------------------------------------GluLeuAsnGlnArg------------GlyLeuAsnCys 1200
          |:::||||||                 ||||||||||
Qy    202 AGCTAAAAGTCACTGTTGGAAACTGGATGGCCACCTTTCAG 163

Db   1200 sLysSerGluArgCysValAsnThrSerGlySerPheArg 1213
          |||                         |||
Qy    163 -----------------------------------GlyLeuUCy 1200
```

US-10-014-896-1 (1-1509) x PCT-US95-02251-3 (1-1251)

```
RESULT 15
US-08-199-780-3

; Sequence 3, Application US/08199780
; Patent No. 5763416
; GENERAL INFORMATION:
;  APPLICANT: Bonadio, Jeffrey
;  APPLICANT: Goldstein, Steven A.
;  TITLE OF INVENTION: Gene Transfer Into Bone Cells
;  TITLE OF INVENTION: And Tissues
;  NUMBER OF SEQUENCES: 3
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: Arnold, White & Durkee
;  STREET: P. O. Box 4433
;  CITY: Houston
;  STATE: TX
;  COUNTRY: USA
;  ZIP: 77210
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/199,780
;  FILING DATE: 18-FEB-1994
;  CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
;  NAME: Parker, David
;  REGISTRATION NUMBER: 32,165
;  REFERENCE/DOCKET NUMBER: UMIC.002
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (512) 320-7200
;  TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 3:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 1252 amino acids
;  TYPE: amino acid
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-08-199-780-3

Alignment Scores:
Pred. No.:
Score:                   0.0739        Length:           1252
                         107.50        Matches:          85
Percent Similarity:      29.97%        Conservative:     31
Best Local Similarity:   21.96%        Mismatches:       133
Query Match:             4.00%         Indels:           139
                         1             Gaps:             20

US-10-014-896-1 (1-1509) x US-08-199-780-3 (1-1252)
```

```
Qy 1158 GACTCTGTTATCAGCCACAAATGTTCTTCGTGAGTTCTAGGACCTTCTTGGACTGTCGTCC 1099
Db  910 AspSerValLeuAlaThrAsnValThrGlnGln                              920

Qy 1098 AGGGTGAATCCGGAAGTTGACTGTGGCCTGGGC---CACTGGGGGGATGACATTGAACTT 1042
Db  921 ----------GluCysCysSerLeuGlyAlaGlyThrTrpGlyAspHisCysGluIle    936

Qy 1041 GACCCCTGCTTGAATATGGTGAGTGC------                              1015
Db  937 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg  956

Qy 1014 ---CGTGGTGGTCCTGATTATTGCAATTGGTTAAGGAGTTTCTCCAATAAACCTGCTTAT   958
Db  957 LeuHisSerGlyGlnHisCys-------------GluLeuCysIleProAlaHis        971
```

```
Db  1190 ----GluLeuAsnGlnArg------GlyLeuLeuCysLysSerGluArgCysValAs 1205
Qy   181 TCTGGAATGGCACCTTCAG                                        163
Db  1205 nThrSerGlySerPheArg                                       1211
```

```
Qy 1158 GACTCTGTTATCAGCCACAAATGTTCTTCGTGAGTTCTAGGACCTTCTTGGACTGTCGTCC 1099
Db  909 AspSerValLeuAlaThrAsnValThrGlnGln                              919

Qy 1098 AGGGTGAATCCGGAAGTTGACTGTGGCCTGGGC---CACTGGGGGGATGACATTGAACTT 1042
Db  920 ----------GluCysCysSerLeuGlyAlaGlyThrTrpGlyAspHisCysGluIle    935

Qy 1041 GACCCCTGCTTGAATATGGTGAGTGC------                              1015
Db  936 TyrProCysProValTyrSerSerAlaGluPheHisSerLeuValProAspGlyLysArg  955

Qy 1014 ---CGTGGTGGTCCTGATTATTGCAATTGGTTAAGGAGTTTCTCCAATAAACCTGCTTAT   958
Db  956 LeuHisSerGlyGlnHisCys-------------                             970

Qy  957 AAGTGGTTCAAATAGCCATGGGGTTGCTCAGGATTATAATTGACAGGGAAGGAAACTCATT 898
Db  971 ArgAspIleAspGluCysIleLeuPheGlyAlaGlyIleCysLysGluGlyLysCysVal  990

Qy  897 TGCCAGTTGCTGCAATACAGTACACCAGCTGTCCCGGCTTCCAAATAT------        853
Db  991 AsnSerGlnProGlyTyrGluCysTyrCysLysGlnGlyTyrTyrAspGlyAsnLeu    1010

Qy  852 ----GATAGGCAATGGTGGTGTCTGCTCCAATCGGCTGCACAGCAGCTGCAAGGA-      806
Db 1011 LeuGluCysValAspValAspGluCysSerLeu                            1027

Qy  805 -----TGCCAAATGCTTGTCTCCTTGGGAGGAGCTG                          776
Db 1028 GlyValCysGluAsnThrArgLeuProCysAlaCys---ThrProProAlaGluTyr   1046

Qy  775 AAGAGTGGCCTGAAGTCAATGTTTACTTCCAGCAATGAGGTTCATGGAACCCTTCTCTGAGA 716
Db 1047 SerProAlaGlnIleAlaGlnCysLeuIle                               1055

Qy  715 CTTGCAATCAAGGCGATGGGCTTCTTGAAGTTAGGAAATGAAAATCATCCAAGAGAGCCCC  656
Db 1056 ----ProGluArgTrpSerThr                                       1061

Qy  655 CCTCGTCCACAATGAAAGGCTAGCTGGACGCCCCTTGACTGTGTAGGCGGTGGTGAGATCCTCT 596
Db 1062 -----ProGln-----ArgAspValLysCysAlaGlyTyrAlaSerGluGlu        1074

Qy  595 GAGCCCCGTCCCCTGATGACTCCTCATCATGGCCCCGAGAAAAATGAAGAAAAGATCTTCGGG 536
Db 1075 ArgThrAlaCysValTrpGlyProThrProAlaGlyPro------AlaLeuThrPheAsp 1091

Qy  535 GGAGGTACTTCCTGATCAGCAGAGACTCCAAGCGCTGCAGTAATGCCATACACAGAGTTCT 476
Db 1092 AspCys-----CysSerArgGlnProArgLeuGly---ThrGln                1103

Qy  475 TGTCGTCTTCCAGTGGTGCCCGACCATAGAATGAACGCCCATCAACCCAGAGAATGGGGG 416
Db 1104 CysSerArgPro---CysProProArgGlyThrGlySerGlnCysProThrSerGlnSerGlu 1122

Qy  415 GCA              CCTCCCAGCCTTCTTCAGGGGCAGGACGCACACATCAAAGTGAG   371
     :::
Db 1123 SerAsnSerPheTrpAspThrSerProLeuLeuLeuGlyLysSerProArgAspGluAsp 1142

Qy  370 ----------CCATCACCA                                           362
Db 1143 SerSerGluGluAspSerGluGluCysValSerGlyProCysValProArgPro      1162

Qy  361 GGTAGGGCTGCAAGCTGGGTCCGAGCCTTGGATAGTGAACAGGTGGCTATACTCTTCCA   302
Db 1163 GlyGlyAlaAlaLysCysGlyCysPro---GlyGlyPheGlnLeuAsp            1176

Qy  301 CGACTTCATGCTTGGATAAAAGCTGGTGGTGACACTGTAGGAAAGAACTTTATGAATGTATT 242
Db 1177 AlaSerArgAla-ArgCysValAspIleAspGluCysArg-                    1189

Qy  241 TTCCGAACTCAGCCAGGCCTGTAGTAATTGGACTTCTCAGAGCTAAAAGTCACTGTTGGAA 182
Db  957 LeuHisSerGlyGlnHisCys-----GluLeuCysIleProAlaHis             971
```

```
Qy   957 AAGTGGTTCAAATAGCCATGGGTTGCTCAGGATTATATTGACAGGAGGAAGGGAAACTCATT 898
             :::              |||          :::|||||||||| :::
Db   972 ArgAspIleAspGluCysIleLeuPheGlyAlaGluIleCysLysGluGlyLysCysVal 991

Qy   897 TGCCAGTTGCTGCAAATACAGTCACCACTGTCCCGCTTCCAAATAT----------- 853
             :::||||       ||||||||
Db   992 AsnSerGlnProGlyTyrGluCysTyrCysLysGlnGlyPheThrTyrAspGlyAsnLeu 1011

Qy   852 -------------GATAGGCATTGGTGTCTGCTCCTCCAAATCGGCTGACAGCACCTGCAAGGA- 806
                        |||:|||                    :::||||:::||:::||:::
Db  1012 LeuGluCysValAspValAspGluCysLeu---------AspGluSerAsnCysArgAsn 1028

Qy   805 ------------------------------TGCCAATGCTTGTCTCCTTTGGAGGAGCTG 853
                                          ||| :::|||
Db  1029 GlyValCysGluAsnThrTrpArgLeuProCysAlaCys---ThrProProAlaGluTyr 1047

Qy   775 AAGAGTGGCCTGAAGTCATGTTTACTTGCACATGAGAGGTTCATGGAACCCTTCTCTGAGA 716
         |||||||||
Db  1048 SerProAlaGlnAlaGlnCysLeuIle---------------------------------- 1056

Qy   715 CTGCAATCAAGGCGATGGGGCTTCTTCTTGAAGTTAGGAATGAAATCATCCAAGATGAAGCCCCC 656
         |||||||::||:::|||
Db  1057 ------------ProGluArgTrpSerThr--------------------------------- 1062

Qy   655 CCTCGTCCACAATGAAAGGCTAGCTGGACGCGCCCCTTGCACTGTTAGCACGGGCTGAGATCCTCT 596
         ||||||||||
Db  1063 -----ProGln--------------------------ArgAspValLysCysAlaGlyAlaSerGluGlu 1075

Qy   595 GAGCCCTGTCCCCTGATGACTCCTCAATCATCAGGCCCAGAGAAATGAGAAAAGATCTTCGGG 536
             :::                              |||||||
Db  1076 ArgThrAlaCysValTrpGlyTyrProTrpGlyPro----------AlaLeuThrPheAsp 1092

Qy   535 GGATGTACTTCCTGATCAGCAGGAGAGCTCCAAGGCCTGCAGTAATGCCATCACAGAGTTCT 476
                                                |||||
Db  1093 AspCys---------CysCysArgGlnProArgLeuGly------------ThrGln 1104

Qy   475 TGTCGTCCAGTGTGCCCCGACCATAGATGACGCCATCACGCTCCAACCCAGAGAATGGGG 416
         |||||||||              |||
Db  1105 CysArgPro--------CysProProArgGlyThrGlySerGlnCysProThrSerGlnSerGlu 1123

Qy   415 GCA---------CCTCCCAGCCTTCTTCAGGGGGCAGGCACCACCATCACAAAAGTGAG 371
         |||                    |||                 ||:::::::||
Db  1124 SerAsnSerPheTrpAspThrSerProLeuLeuLeuGlyLysSerProArgAspGluAsp 1143

Qy   370 -----------------------------------------CCATCAGCA 362
                                                  |||
Db  1144 SerSerGluGluAspSerGlyArgCysValSerGlyProCysValProArgPro 1163

Qy   361 GGTAGGGCTGCAAGCTGGGGTGGTCCGAGCCTTGGATAGTGAACAGGTGGCTATACTCTTCCA 302
         |||               ||| 
Db  1164 GlyGlyAlaValCysGluCysPro----------GlyGlyPheGlnLeuAsp 1177

Qy   301 CGACTTCATGCTGCTGGATAAAGCTGGTGGTGCTGACCACTGTAGGAAAGACTTTATGAAGTATT 242
         |||
Db  1178 AlaSerArgAla-ArgCysValAspIleAspGlyCysArg--------------- 1190

Qy   241 TTCCGAACTCACCCAGGGCCTGTAGTATTGGACTTCTCAGAGCTAAAAGTCACTGTTGGAA 182
                                                |||||:::|||
Db  1191 ---------GluLeuAsnGlnArg--------GlyLeuLeuCysLysSerGlnUargCysValAs 1206

Qy   181 TCTGGATGGCACCTTTCAG 163
         |||:::::|||:::
Db  1206 nThrSerGlySerPheArg 1212
```

```
OM nucleic - protein search, using frame_plus_n2p model

                    GenCore version 5.1.6
                    Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:             June 27, 2003, 11:03:29 ; Search time 69 Seconds
                                               (without alignments)
                                               4795.857 Million cell updates/sec

Title:              US-10-014-896-1
Perfect score:      2762
Sequence:           1 atggctcaagcggtgcgtttg.........ctcacctgcacaaactgtga 1509

Scoring table:      BLOSUM62
                    Xgapop  10.0 ; Xgapext  0.5
                    Ygapop  10.0 ; Ygapext  0.5
                    Fgapop   6.0 ; Fgapext  7.0
                    Delop    6.0 ; Delext   7.0

Searched:           424699 seqs, 109646833 residues

Total number of hits satisfying chosen parameters:       849398

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_n2p.model -DEV=rlp
-Q=/cgn2_1/USPTO_spool/US10014896/runat_27062003_104417_10539/app_query.fasta.1.1671
-DB=Published_Applications_AA -QFMT=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=-1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DGCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US10014896 @CGN_1_1_18 @runat_27062003_104417_10539
-NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
-LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
-FGAPOP=6 -FGAPEXT=7 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA *
  1 /cgn2_6/ptdata/1/pubpaa/US08_NEW_PUB.pep:*
  2 /cgn2_6/ptdata/1/pubpaa/US08_NEW_PUB.pep:*
  3 /cgn2_6/ptdata/1/pubpaa/PCT_NEW_PUB.pep:*
  4 /cgn2_6/ptdata/1/pubpaa/US06_NEW_PUB.pep:*
  5 /cgn2_6/ptdata/1/pubpaa/US06_PUBCOMB.pep:*
  6 /cgn2_6/ptdata/1/pubpaa/US07_NEW_PUB.pep:*
  7 /cgn2_6/ptdata/1/pubpaa/US07_PUBCOMB.pep:*
  8 /cgn2_6/ptdata/1/pubpaa/PCTUS_PUBCOMB.pep:*
  9 /cgn2_6/ptdata/1/pubpaa/US09_NEW_PUB.pep:*
 10 /cgn2_6/ptdata/1/pubpaa/US09_PUBCOMB.pep:*
 11 /cgn2_6/ptdata/1/pubpaa/US10_NEW_PUB.pep:*
 12 /cgn2_6/ptdata/1/pubpaa/US10_PUBCOMB.pep:*
 13 /cgn2_6/ptdata/1/pubpaa/US60_NEW_PUB.pep:*
 14 /cgn2_6/ptdata/1/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

```
Result                    %
No.   Score  Query   Match Length  DB   ID            Description
----- ------ ------- ----- ------  ---- -----         -----------
1     275.5  10.0    441   9       US-09-738-626-6613  Sequence 6613, Ap
2     190    6.9     773   12      US-10-109-860-2     Sequence 2, Appli
3     185.5  6.7     408   12      US-10-109-860-4     Sequence 4, Appli
4     151.5  5.5     476   9       US-10-073-885-103   Sequence 103, App
```

(handwritten: Nothing to cite)

ALIGNMENTS

RESULT 1
US-09-738-626-6613
; Sequence 6613, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: Patentin ver. 3.0
; SEQ ID NO 6613

```
  5  151.5  5.5  476  10  US-09-815-242-12290  Sequence 12290, A
  6  149.5  5.5  286   9  US-10-073-885-78     Sequence 78, Appl
  7  146.5  5.3  403  10  US-09-815-242-10282  Sequence 10282, A
  8  145    5.2  449  10  US-09-815-242-5375   Sequence 5375, Ap
  9  145    5.2  469  10  US-09-815-242-12290  Sequence 12290, A
 10  136.5  4.9  457   9  US-09-738-626-6496   Sequence 6496, Ap
 11  136.5  4.9  501   9  US-09-791-389-2      Sequence 2, Appli
 12  136.5  4.9  501   9  US-09-791-393-2      Sequence 2, Appli
 13  136.5  4.9  501  10  US-09-791-378-674    Sequence 674, App
 14  136.5  4.9  507   9  US-10-033-041-57     Sequence 57, Appl
 15  136.5  4.9  507   9  US-10-035-085-57     Sequence 57, Appl
 16  136.5  4.9  507   9  US-09-931-836-57     Sequence 57, Appl
 17  136.5  4.9  507   9  US-10-036-214-57     Sequence 57, Appl
 18  136.5  4.9  507   9  US-09-815-242-5877   Sequence 5877, Ap
 19  136.5  4.9  507   9  US-10-035-719-57     Sequence 57, Appl
 20  136.5  4.9  507   9  US-10-036-160-57     Sequence 57, Appl
 21  136.5  4.9  507   9  US-10-033-958-57     Sequence 57, Appl
 22  136.5  4.9  507   9  US-10-036-150-57     Sequence 57, Appl
 23  136.5  4.9  507   9  US-10-036-063-57     Sequence 57, Appl
 24  136.5  4.9  507  12  US-09-963-290-2      Sequence 2, Appli
 25  136.5  4.9  508   9  US-10-036-342-57     Sequence 57, Appl
 26  136.5  4.9  508   9  US-09-948-783-139    Sequence 139, App
 27  136.5  4.9  509  10  US-09-731-872-242    Sequence 242, App
 28  133.5  4.8  471   9  US-09-892-877-137    Sequence 137, App
 29  132    4.8  509  10  US-09-815-242-10682  Sequence 10682, A
 30  132    4.8  466  10  US-09-815-242-13394  Sequence 13394, A
 31  123    4.5  190  10  US-09-815-242-13537  Sequence 13537, A
 32  123    4.5  215  10  US-09-815-242-5877   Sequence 5877, Ap
 33  119    4.3  478   9  US-09-738-626-3883   Sequence 3883, A
 34  118    4.3  2462  9  US-09-948-783-139    Sequence 139, App
 35  114    4.1  890   9  US-09-731-872-242    Sequence 242, App
 36  113.5  4.1  412  10  US-09-373-658-4      Sequence 4, Appl1
 37  113.5  4.1  412  10  US-10-289-360-4      Sequence 4, Appl1
 38  113    4.1  1284 10  US-09-950-772-6      Sequence 6, Appl1
 39  113    4.1  1284  9  US-09-160-224-9      Sequence 9, Appl1
 40  113    4.1  1284 10  US-10-185-867-11     Sequence 11, Appl
 41  111    4.0  5877  9  US-09-119-855-11     Sequence 11, Appl
 42  111    4.0  383   9  US-10-142-515-11     Sequence 11, Appl
 43  110.5  4.0  383   9  US-10-285-074-3      Sequence 3, Appl1
 44  107    3.9  420  10  US-09-727-238-4      Sequence 4, Appl1
 45  105    3.8  5935 10  US-10-243-243A-8     Sequence 8, Appl1
 --  104    3.8  2586 10  US-09-905-129-11     Sequence 11, Appl1
```

```
; LENGTH: 441
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-6613

Alignment Scores:
Pred. No.:
Score:              4.03e-16
Percent Similarity: 275.50
Best Local Similarity: 41.18%
Query Match:        24.79%
DB:                 9.97%

                    Length:       441
                    Matches:      118
                    Conservative: 78
                    Mismatches:   195
                    Indels:       85
                    Gaps:         23
```

```
US-10-014-896-1 (1-1509) x US-09-738-626-6613 (1-441)

Qy   136  GAGGAACGCGTCGCGATAAAGAGGAGGGCTGCC----ATCCAGATTCCAACAGTG  192
          |||                |||||||||||||||        :::   ||||||||
Db    6   GluThrLeuThrLeuLeuGlnGluLeuIleArgAsnAlaCysValAsnAspLeuThrPro  25

Qy   193  ACTTTTTAGCTCTGAGAAGTCAAATACTACAGCCCTGGCTGAGTTCGGAAAATACATTCAT  252
          ::::||                    :::||||||||||||||||||||
Db    26  AspSerGlyIleGluIleIleArgAsnAlaGluSerLeuGluArgPhe----------Phe  41

Qy   253  AAAGTCTTTTCCTACACAGTGGTCAGCACCAGCTTTATCCCAGATCGTGGAAGAGTAT  312
          |||   :::  |||||||||||||||||||||||||||||||||
Db    42  GluGlyThrProAsnValLysIleThrLysLeuGluProHisPro-------GlyArgThr  59

Qy   313  AGCCACCTGTTCACTATCCAAAGCTCGGACCCCATGGTTGCAAGCCCTACCTTCTCGATGGCT  372
          |||   :::  |||||||||||||||||||||||||||
Db    60  SerIleIleValThrValProGlySerAspSerProAlaGluGlnLeuGlyLeuGly  79

Qy   373  CACTTTGATGTGGTTGGTGCCTGCCCCCTGGAAGAGGCTGTGGGAGGTGCCCCCATTCTCTGGGGTTG  432
          |||   |||||||||||||
Db    80  HisThrAspValValProGlySerAspLeuProLysTrpThrLysProPheGlyAlaGlu  99

Qy   433  GAGGCGTGAATGGGGTGTCACTCTATGGTCGGCGGACACTGACGACAGAACTCTGTGATGGCA  492
          |||||||||||||||||||||||||||:::::||||
Db    100  IleSerAspGlyGlnIleThrProLyArgGlySerValAsp------------MetLeu  114

Qy   493  TTACTGCAGGCCTTGGAGCTCCTGCCTGACTCAGGAGTACATCCCCGGAAGATCT-----  546
          |||   |||||||||||||||||||||||      |||
Db    115  PheIleThrAlaThrGlnAlaAlaValThrArgGln---ValAlaArgGluGlyLeu  133

Qy   547  -----------TTCTTCATTTCTCTGGGCCATGAGTGAGGAGTCATCAGGAGCAGGGGCT  594
                       ::  |||       |||||||||
Db    134  ArgGlyThrLeuThrPheValGlyValAlaAspGluGluAlaArgGlyValGlyLeuValAla  153

Qy   595  CAGAGGATCTCA----------------GCCCTGCTACAGTCA  621
          :::                        |||
Db    154  LysTrpLeuSerGluGluHisGlnAsnLeuPheSerTrpLysAsnCysLeuSerGluSer  173

Qy   622  AGGGGCGGTCCAGCTAGCCTTCATTGTGGACAGGGGGCTTCATCTTGGATGATTTCATT  681
          |||
Db    174  GlyGlySerHisLeuPro-----ValHisAspGly  183

Qy   682  CCTAACTTCAAGAAGCCCATCGCCTTGATTGCAGTCTCAGAGAAGGGTTCACATGAACCTC  741
          :::|||   |||
Db    184  ------------SerAspAlaValValIleAsnValGlyGluLysGlyAlaAlaGlnArg  199
```

```
Qy   742  ATGCTGCAAGTAAAACAATGACTTCAGGCCCTCAGCCTCCT---CCAAAGAGAGCAAGC  798
          ::::::||||||||||          :::::::::
Db    200  ArgIleHisValAsnGlyAspAlaGlyHisGlySerIleProPheProAspSerAla  219

Qy   799  ATTGGCATCCTTGCACCTGTCTCAGCCGATTGGAGCAGAGAACCAATGCCTATCATA---  855
          |||                |||
Db    220  IleValLysIleGlyValAlaAlaArgAlaIleAlaAlaAspLeuLysValAlaLys  239

Qy   856  ------TTTGGAAGCGGACAAGTGGTGACTGTTATTGCAGCAACTGGCAAATGAGTTT--  906
          :::
Db    240  AspAspIleTrpGlnGlyPheVal-------------GlnAlaHisArgPheAsp  253
```

```
Db    254  ProGluThrGluGlnAlaLeuLeuSerGlyThrSerProGluAlaThrTyrAlaGluPheGly  273

Qy   958  ---ATAAGCAGGTTTATGGAGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACGACCACACG  1014
             :::|||||||||||
Db    274  GlyLeuSerArgPhe----------------AlaHisSerHisLeuThrIle  287

Qy   1015  GCACTCACCATATTCAAAGCACAGGGTAACAGAGGGGTCAAGTTCAATGTCATCCCCCAGTTGGCCCAGCC  1074
           |||||||||||                   :::||||||||||||||
Db    288  AlaGlnThrValValArgAlaGlyGlnAlaIleAsnValLeuProSerHisAlaTyrLeu  307

Qy   1075  ACAGTCAACTTCCGGATTCACCCTGGGACAGACA------GTCCAAGAGGTCCTAGAA  1125
           |||||||||                  :::||||||
Db    308  GluLeuAspIleArgPheLeuProGlyThrAsnAspThrValValAspAspThrLeuArg  327

Qy   1126  CTCACGGAAGAACATATGTGGCCTGATAAACAGAGTCCAGTTCCATGTGTTGTTGAGTGCCTTTGAC  1185
           |||||||||||||||||||          ||||||||
Db    328  AlaAlaLeuGlyAspLeuAlaAspGluValGluIleGluHisLeuIleSer------Glu  345

Qy   1186  CCCCTCCCGTCACGCCTTCTGATGACAAGGGCCTTGGGGTAGCTTCAGCTGCCTCCGGAGACC  1245
           |||||||||||||||||||          :::||||||
Db    346  GluAlaThrValSerProThrArgAspSerArgLeu---------TyrAsnThrGluGlyLysVal  363

Qy   1246  GTACAGTCCGTCTTCCCGGAAGTCAATATTAACTACCTCCCCAGTTGTTCTTATTGGCACAA  1305
           |||||||||||          :::||||||
Db    364  LeuGlyAspPhePheProAspSerAlaProVal---ValProIleIleSerGlyGlySer  382

Qy   1306  GACAGCCGATTCTTTACAAACCTCACCACTGGCATCTACAGGTTCTACCCCATCACTA  1365
           |||||||||||||          :::||||||
Db    383  AspLeuArgPheGlyArgArgLeu-------ValProIleIleSerGlyGluAlaHisAla  401

Qy   1366  CAG-----------CCTGAAGACTTCAAACGACCATCATGGAGTCAACGAG  1404
           :::                |||
Db    402  ArgGlyLeuArgThrLeuAlaGluAlaMetGlyGlnLeuHisSerHisAspGluAlaLeuTyr  421

Qy   1405  ------------AAAATCTCAGTCCAAGCCTATGAGACCCCAAGTGAA  1440
                       :::||||||
Db    422  LeuGluAspLeuGluLeuThrValArgGlyThrArgAspSerValValArg  437
```

RESULT 3
US-10-109-860-4
Sequence 4, Application US/10109860
Patent No. US20020142421A1
GENERAL INFORMATION:
   APPLICANT: SHAO, Wei et al.
   TITLE OF INVENTION: ISOLATED HUMAN AMINOACYLASE, NUCLEIC
   TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN AMINOACYLASE, AND USES THEREOF
   FILE REFERENCE: CL001179DIV
   CURRENT APPLICATION NUMBER: US/10/109,860
   CURRENT FILING DATE: 2002-04-01
   PRIOR APPLICATION NUMBER: 09/814,951
   PRIOR FILING DATE: 2001-03-23
   NUMBER OF SEQ ID NOS: 14
   SOFTWARE: FastSEQ for Windows Version 4.0
   SEQ ID NO 4
   LENGTH: 408
   TYPE: PRT
   ORGANISM: Homo sapiens
US-10-109-860-4

DB:

Alignment Scores:
Pred. No.:
Score: 5.46e-08   Length: 408
Percent Similarity: 185.50   Matches: 89
Best Local Similarity: 39.14%   Conservative: 57
Query Match: 23.86%   Mismatches: 158
                               6.72%   Indels: 69
                               12   Gaps: 13

US-10-014-896-1 (1-1509) x US-10-109-860-4 (1-408)

[The remainder of the page consists of a paired nucleotide/protein sequence alignment, arranged in two columns of Qy (query) and Db (database) rows, e.g.:]

Qy   352 CAGCCCTACCTGCTGATGGCTCACTTTGATGGGTGGTGCCCCGTCCCCTGAACAAGGCTGGGAG 411
Db   38 SerSerIleLeuLeuAsnSerHisThrArgProValValProValPheLysGluHisTrpSer 57

Qy   412 GTGCCCCCAATTCTCTGGGGTTG---GAGGCTGAATGGCGTCATCTATGGTGGGGCGCACACTG 468
Db   58 HisAspProPheGluAlaPheLysAspSerGlyGlyTyrIleTyrAlaArgGlyAlaGln 77

Qy   469 GACGAACGAAACTTGTGATGATTACTGCAGGGCCATTACTGCAGGCGTCCTGCTGATCAGGGAG 528
Db   78 AspMetLysCysValSerIleGlnThrTyrLeuGlnAlaIleAlaValArgArgLeuLysValGlyGly 97

Qy   529 TACATCCCCGAAGATCTTTCTTCATTTCTCTGGGCCCCATGATGAGGAGTCATCAGGGACA 588
Db   98 HisArgPheProArgThrIleHisPheValProAspGluValGlyGly 116

Qy   589 GGGGCTCAGGAGGATCTCAGCCCCTGCTACAGTCAAGGGGGTCCACGTA--- 636
Db   117 ---GCCTTCATTGTGGGACGAGGGGGCTTCATCTGGGATGAATTCATTCCTAACTTCAAG 126

Qy   694 AAGCCCCATCGCCTTGATTGCAGTC------TCAGAGAAGGGGTTCCATGAACCTCATGCTG 747
Db   145 ---ProThrAspAlaPheThrValPheTyrSerGluArgSerProTrpProValArgVal 163

Qy   748 CAAGTAAACACATGACTTCAGGCCACTGTTCAGCTCTCCCCAAAGGAGACAACAGGCCATTGGCCATC 807
Db   164 ThrSerThrGlyArgProGlyHisAlaSerArgPheMetGluAspThr 179

Qy   808 CTTGCAGCTCGTGCAGCCCGATTGGAGCAGCAACAATGCCTACTATATTTGGAAGCGGG 867
Db   180 ---AlaAlaGluLysLeuHisLys- 186

Qy   868 ACAGTGGTTGGACTGTATTGCAGCAGTGGCAAATGAGTTTCCGTTCCCTGTCAATATAATC 927
Db   187 ValValAsnSerIleLeuAlaPheArgGluLysGluTrp- 202

Qy   928 CTGAGCAACCCATTGGCTATTTGAACCACTTTATAAGCAGGTTTATGGAGAGAAATCCCTTA 987
Db   203 GlnSerAsnProHisLeuLysGlyLySerVal- 213

Qy   988 ACCAAATGCAATAAATCAGGACCACCACGTCATCAGCATAATTCAAAGCACAGGGGTCAAGTTC 1047
Db   214 ---ThrSerValAsnLeuThrLysLeuGlyIleValAlaThr 227

Qy 1048 AAATGTCATCCCCCAGTGGCCCAGAACGTACAGTGGCCCTCCGGGATTCACCCTGGACAGACA 1107
Db   228 AsnValIleProAlaThrMetSerAlaSerPheAspPheArgValAlaProAspValAsp 247

Qy 1108 GTCCAAGAGGTGCCTAGAACTCACGAAGAGAAC------ATTGTGGCTGATAACAGAGTGCCAG 1161
Db   248 PheLysAlaPheGluLeuGlnLeuGlnSerArgTrpCysGlnAlaAlaAlaGlyGlyValThr 267

Qy 1162 TTCCAGTGTTGAGTGCCTTTGACCCCTCCCGGTCCAGCCCTTCTGATGACAAGGCCTTG 1221
Db   268 LeuGluPheAlaAlaGlnLysTrpMetHisProGlnValThrProThrAspAspSerAsnPro 287

Qy 1222 GGCTACCAGCGTGCTCCCAGCACGTACAGTCCGTGTTCCCGGAAGTCAATATTACT--- 1278
Db   288 TrpTrpAlaAlaPheSerArg- 303

Qy 1279 GCCCCAGTTACTTCTTCTTATTGGCAACACAGCACCACCGATTCTTTAAAACCTACCACTGGC 1338
Db   304 GluProGlyIleLeuMetProAlaAlaThrAspAspAsnArgTyrIleArgAlaValGlyValPro 323

Qy 1339 AATCTACAGGTTCTACCCCATCTACATACAGCCT------GAAGACTTCAAACGC 1386
Db   324 AlaLeuGlyPheSerProMetAsnArgThrProValLeuLeuHisIleAspGluArg 343

Qy 1387 ATCCAT 1392
Db   344 LeuHis 345

Qy   319 CTGTTCACTATCCAAGGCTCGGACCCCCAGCTTGCACCCTACCTGCTGATGGCTCACTTT 378
Db   62 ValLeuThrThrProProGlyThrAsnProThrHisLeuSerSerIleLeuLeuAsnSerHisThr 81

Qy   379 GATGGTGGTGCCTGCCCTGCCCCCTGAAGAAAGGCTGGGAGGTGAGGGGTGCCCCCAATTCTCTGGGGTTG---GAG 435
Db   82 AspValValProValPheLysAspSerGlyGlyTyrIleTyrAlaArgGlyAlaGlnAspMetLysCysValSerIleGlnThrTyr 101

Qy   436 CGTGATGGCGTCATCTATGGTGGGGGCGCACACTGGACGACAAGAACTCTGTGATGGCCATTA 495
Db   102 SerGluGlyTyrIleTyrTyrAlaArgGlyAlaGlnAspMetLysCysValSerIleGlnThr 121

Qy   496 CTGCAGGCGCCTTGGAGCTCCTCCGTCGATCAGGAAGTACATCCCCGAAGAGATCTTCATT 555
Db   122 LeuGluGluLysTrpMetHisProGlnValThrProThrAspAspSerAsnPro 141

Qy   556 TCTCTGGGCCCATGATGAGGAGTCATCAGGGACAGGGGCTCAGGAGGATCTCAGCCCCTGCTA 615
Db   142 ThrPheValProAspGluValGlyGly- 151

Qy   616 CAGTCAAGGGGGGTCCAGCTA------GCCTTCATTGTGGGACGAGGGGGC 660
Db   152 ---HisGlnGlyMetGluLeuPheValGlnArgProGluPheHisAlaLeuHisArgAlaGly 170

Qy   661 TTCATCTGGGATGAATTCATTCCTAACTTCAAGAAGAGCCCAGCCGTTGCAGTC--- 717
Db   171 PheAlaLeuAsnProGluGlyIleLeuAlaAsn- 187

Qy   718 ---TCAGAGAAGGGGTTCCATGAACCTCATGCTGGCTACTGGTCTGCTCAGTAAACATGACTTCAGGCCACTCT 774
Db   188 TyrSerGluArgSerProTrpProValArgGlyHisAla 207

Qy   775 TCAGCTCTCCCAAAGGAGACAACAGGCCATCCTTGCCAGCTGCCTGATGGAG 834
Db   208 SerArgPheMetGluAspThr- 220

Qy   835 CACACCCAATGCCTACTATCATATTTTGGAAGCGCGGACACTGGTGGTCACTGTATTCCAGCAACTG 894
Db   221 Lys------ 230

Qy   221 ValValAsnSerIleLeuAlaPheArg 230

RESULT 4

US-10-073-885-103

Sequence 103, Application US/10073885
Publication No. US20030096346A1
GENERAL INFORMATION:
  APPLICANT: Rosen et al.
  TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
  FILE REFERENCE: PJZ03C1
  CURRENT APPLICATION NUMBER: US/10/073,885
  CURRENT FILING DATE: 2002-02-14
  PRIOR APPLICATION NUMBER: See file Wrapper or Palm
  NUMBER OF SEQ ID NOS: 116
  SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 103
  LENGTH: 476
  TYPE: PRT
  ORGANISM: Homo sapiens
  FEATURE:
  NAME/KEY: misc_feature
  LOCATION: (127)
  OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-10-073-885-103

Alignment Scores:
  Pred. No.:
  Score: 7.02e-05
  Percent Similarity: 151.50
  Best Local Similarity: 35.96%
  Query Match: 21.58%
  DB: 5.498
                            9

                            Length: 476
                            Matches: 93
                            Conservative: 62
                            Mismatches: 183
                            Indels: 93
                            Gaps: 17

US-10-014-896-1 (1-1509) x US-10-073-885-103 (1-476)

Qy   1330 ACCACTGGCAATCTACAGGTTCTTACCCCATCTACATACAGCCTGAAGACTTCAAACGCCATC 1389
          ||| :|| ::|||
Db   429 ThrGlyLysAsnValMetLeuLeuProValGlySerAlaAspsGly------------Ala 445

Qy   1390 CATGGAGGTCAACGAGAGAAAATCTCA-----------------GTCCAAGCCTATAGAGACCCAAGTG 1437
          |||||||||||||||||||||||||| :|||
Db   446 HisSerGlnAsnGluLysLeuAsnArgTyrAsnTyrIleGluGlyThrLysMetLeuAla 465

Qy   1438 AAATTCATCTTTGAGTTTGATTCAGAATGCTGAC 1470
          ::|||:||||:||||:||:| :|||
Db   466 AlaTyrLeuTyrGluValSerGlnLeuLysAsp 476

RESULT 5
US-09-925-301-1397
; Sequence 1397, Application US/09925301
; Patent No. US20052008A1
; GENERAL INFORMATION:
;   APPLICANT: Rosen et al.
;   TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
;   FILE REFERENCE: PA106
;   CURRENT APPLICATION NUMBER: US/09/925,301
;   CURRENT FILING DATE: 2001-08-10
;   PRIOR APPLICATION NUMBER: PCT/US00/05882
;   PRIOR FILING DATE: 2000-03-08
;   PRIOR APPLICATION NUMBER: 60/124,270
;   PRIOR FILING DATE: 1999-03-12
;   NUMBER OF SEQ ID NOS: 1694
;   SOFTWARE: Patentin Ver. 2.0
;   SEQ ID NO 1397
;   LENGTH: 476
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: SITE
;   LOCATION: (127)
;   OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-301-1397

Alignment Scores:
Pred. No.:
Score: 7.02e-05
Percent Similarity: 35.96%
Best Local Similarity: 21.58%
Query Match: 5.498
DB: 10

Length: 476
Matches: 93
Conservative: 62
Mismatches: 183
Indels: 93
Gaps: 17

US-10-014-896-1 (1-1509) x US-09-925-301-1397 (1-476)

Qy   334 GGCTCGGACCCCAGCTGCAGCCCTACCTGCTGAGGGCTCACTTTGATGTGGTGCCTGCC 393
          :|||||||||:||||||:|
Db   87 GlySerAspProGlnLysThrValCysIleTyrHisLeuAspValGlnProAla 106

Qy   394 CCT---GAAGAAGGCTGGGGAGGTGGCCCCCCATTCTCTGGGGTTGGAGGCGGTGGTCATC 450
          ||:||::||||||||||||
Db   107 AlaLeuGluAspGlyTrpProPheThrLeuValGluArgAspGlyLysLeu 126

Qy   451 TATGGTCGGGGCCACACTGGACGACAAGAACTCTGTGATGGCCATTACTGCAGGCCTTGGAG 510
          |||||||||:::
Db   127 ***GlyArgGlySerThrAspSerProValGlyProValAlaGlyTrpIleAsnAlaLeuGlu 146

Qy   511 CTCCTG---CTGATCAGGAAGTACATCCCCGAGATCTTTTCTTCATTTCTCTGGGCCAT 567
          |||||| :|||
Db   147 AlaTyrGlnLysThrGlyGlnIleLeuProValAlaArgPheCysLeuGluGlyMet 166

Qy   568 GATGAGGAGTCATCAGGGACAGGGGCTCAGAGGATCCAGCCCTACTACGTCAGTCAAGGGGC 627
          :|| |||||||
Db   167 GluGlu------------SerGlySerGluGlyLeuAspGluLeuIlePheAlaArg--- 181

Qy   628 GTCCAGCTAGCCTTCATTGTGGAC---GAGGGGGGCCTTCAATCTTGGATGATTTCANTTCCT 684
          ::| |||
Db   182 ---LysAspThrPhePheLysAspVAlValCysIleSerAspAsnTyrTrpLeu 200

Qy   685 AACTTCAAGAAGCCCATCGGCCTTGATTGCAGTCTCAGAGAAGAAGGGTTCCAATGAACTCANTG 744

; Prior Application removed - See file Wrapper or Palm
; SOFTWARE: Patentin Ver. 2.0
; NUMBER OF SEQ ID NOS: 116
; SEQ ID NO 78
; TYPE: PRT
; LENGTH: 286
; ORGANISM: Homo sapiens
US-10-073-885-78

Alignment Scores:
Pred. No.:
Score: 8.39e-05
Percent Similarity: 149.50
Best Local Similarity: 46.93%
Query Match: 29.05%
DB: 5.41%

Length: 286
Matches: 52
Conservative: 32
Mismatches: 76
Indels: 19
Gaps: 8

US-10-014-896-1 (1-1509) x US-10-073-885-78 (1-286)

```
Qy   337 TCGGACCCCAGCTTGCAGCCTACCTCGATGGCTCACTTTGATGTGGTGGCCTGCCCCT 396
             ||||||||      ||||        ||| |    |||||||||| |  |||||||
Db    43 SerAspProGlnLysLysThrValCysIleTyrGlyHisLeuAspValGlnProAlaAla 62

Qy   397 --GAAGAAGGCTGGGGAGGTGCCCCCATTCTCTGGGTTGGAGGCGGTGGATTCTAT 453
             ||||||||:||||:||                ||| ::||
Db    63 LeuGluAspGlyTrpAspSerGluProPheThrLeuValGluArgAspGlyLysLeuTyr 82

Qy   454 GGTCGGGGCACACTGACGACAAGAAACTCTGTGATGGCATTACTGCAGGCCTTGGAGCTC 513
             ||||:|||||| :||||||||||||||         ||  :||
Db    83 GlyArgGlySerThrAspAspLysLysGlyProValAlaGlyLeuLeuGlnAlaLeuGluAla 102

Qy   514 CTG---CTGATCAGGAAGTACANCCCCGAGAGATCTTTCTTATTTTCTGGGCCATGAT 570
             |||      ||         :||                      |||||||||:||
Db   103 TyrGlnLysThrGlyGlnIleProValAsnValArgPheCysLeuGlyLysMetGlu 122

Qy   571 GAGGAGGTCATCAGGGACAGAGGGCTCAGAGGGATCTCAGCCCCTACAGTCAAGGGGGTC 630
             ||||||:|||||     ||:|| |||  ||||
Db   123 Glu--------SerGlySerGluGlyLeuAspGluLeuPheAlaArg------ 136

Qy   631 CAGCTAGCCTTCANTTGTGGAC---GAGGGGGGCTTCANCTGGATGANTTCANTCCTAAC 687
             |||   :|||||  :||     |||||||||||    ||||||:||||||:||:|||||:
Db   137 LysAspThrPhePheLysAspValCysIleSerSerAspAsnTyrTrpLeuGly 156

Qy   688 TTCAAGAAGGCCCANCGCCTTGACTCTCAGAGAAGGGGTTCCANGAACCTCATGCTG 747
             ||||||||||      || ||      ||          ::|     :||:
Db   157 LysLysLysPro-----CysIleThrTyrGlyLeuArgGlyTyrLeuCysTyrPhePheIle 174

Qy   748 CAAGTAAACATACANCTTCAGGCCACTCTTCAGCTCCTCCAAAGGAGACAAAGCANTTGGCATC 807
             :||| ||     :|||          :||||          ::||
Db   175 GlyValGluCysSerAsn--------- 187

Qy   808 CTTGCACAGCTGCTTCACCGATTGGACACCAACATGCCTTATCATATATTTGGAAGC 864
             ::||| ||            ||| :|:||                  ||||
Db   188 TyrGlyGlySerValHisGlu---AlaMetThrAspIleLeuLeuMetGlySer 205
```

; Prior Application removed - See file Wrapper or Palm
; SOFTWARE: FastSEQ for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 14110
; SEQ ID NO 10282
; TYPE: PRT
; LENGTH: 403
; ORGANISM: Escherichia coli
US-09-815-242-10282

Alignment Scores:
Pred. No.:
Score: 0.000184
Percent Similarity: 146.50
Best Local Similarity: 34.89%
Query Match: 18.44%
DB: 5.30%

Length: 403
Matches: 83
Conservative: 74
Mismatches: 166
Indels: 127
Gaps: 16

US-10-014-896-1 (1-1509) x US-09-815-242-10282 (1-403)

```
Qy   106 GCGGTNCGGCGAATCCCTTCTTCAGTTC-----------AGCAAAGAGGAACGGCTCGGGATG 153
             |||   :|||              ||                ::
Db     2 AlaLysAsnIleProPheLysLeuIleLeuGluLysAlaAlaLysAspArgGlnAlaAsnAlaAlaMet 21

Qy   154 AAAGAGCGCTGAAAAGGTGCCATCCAGAATTCCAAACAGTGACTTTTAGCTCTGAGAAGTCC 213
             |||||||:|:||          |||:|||||||||:||||:|||:|||||||||:||||||
Db    22 ThrArgPheLeuArgAspMetValAlaIleAlaIleProSerGluSer---CysAspGlyLys--- 39

Qy   214 AATACTACAGCCCTGGCTGAGTTCGGAAAAATACATTCATAAAGTCTTTCCTACAGTGGTC 273
             ||||||||||    |||||||:|||||||   |||||||||||    :||:
Db    40 -----------ArgValAlaHisArgIleLysGluLysGluMetGlu 50

Qy   274 AGCACCAGCTTTATCCAGACNCATGAAAGGTCGTGGAAGAGTATAGCCACCTGGTTCACTATCCAA 333
             ||||||||||||||:|||    |||||||:||:|||||||||:|||||||||:|||:||||||
Db    51 LysLysValGlyPheAspLysValGlu---IleAspProMetGlyAsnValLeuGlyTyrIle 69

Qy   334 GGGCTCGGGACCCCAGCTTGCAGCCCTACCTGATGGCTCACTTTGANGTGGTCGGCCTGCC 393
             |||||||||||||||||||||||:|||||||||||||||||||||:||||:|||:|||||
Db    70 GlyHisGlyProArgLeu-----ValAlaMetAspAlaHisIleLeuAspThrValGlyIle 87

Qy   394 CCTGAA--GAAGGCTGGGGAGGTGCCCCCCATTCTCTGGGTTGGAGGCGGTGATGGGTCATC 450
             ||||        ||||||||||||||||:|||||||||||||||||||||:|||||||
Db    88 GlyAsnIleLeuLysAsnTrpAspPheAspProTyrGluGlyMetGluThrAspGluLeuIle 107

Qy   451 TANGGTCGGGGCACACTGACGACAAGAAACTCTGTGATGGCCATTACTGCAGGCCTTGGAG 510
             |:||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db   108 GlyGlyArgGlyThrSerAspGlnGluGlyIleGlyMetAlaSerMetValThrValAlaGlyLys 127

Qy   511 CTCCTGCTGATCAGGAAGTACANTCCCCGAGAGATCTTTCTTCATTTTCTGGGCCATGAT 570
             |||:|||          |:|:||                ::||
Db   128 IleLeuGlyLys----------------------AspLeuGlu 135

Qy   571 GAGGAG-------TCATCAGGGACAGAGGGCTCAGAGGGATCTCAGCCCCTGCTACAG 618
             ||||||          :|||                 :|||:|:|
Db   136 AspGluTyrThrLeuLeuValAlaThrGlyThrValGlnGluGlu--- 150

Qy   619 TCAAGGGGGTCCAGCTAGCCTTCANTTGTGGACGAGGGGGGCTTCANCTGGATGANTTTC 678
             |||:|||||||||:|||     ||  :||     |||||||||||  ::|||||||:|||
Db   151 CysAspGlyLeuCysTrpProGlnIleTyrIleIleGluIleGlnSerGly---IleArgProGluPhe 169

Qy   679 ANTCCTAACTTCAAGAAGGCCCANCGCCTTGATTGCAGTCTCAGAGAGAAGGGTTCCATGAAC 738
             |:||| || ||||||||||      || ||      ||          ::|
Qy
```

GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
    Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21

RESULT 7
US-09-815-242-10282
Sequence 10282, Application US/09815242
Patent No. US2002001569A1

PRIOR APPLICATION NUMBER: 60/191,078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206,848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207,727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242,578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253,625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2000-12-22
PRIOR APPLICATION NUMBER: 60/269,308
PRIOR FILING DATE: 2001-02-16
PRIOR APPLICATION NUMBER: 60/257,931
PRIOR FILING DATE: 2001-02-16

RESULT 8
US-09-815-242-5375
Sequence 5375, Application US/09815242
Patent No. US20020061569A1

DB:

Query Match:
Best Local Similarity: 21.67%
Percent Similarity: 35.00%
Score: 145.00
Pred. No.:   0.000265

Alignment Scores:

Length: 449
Matches: 104
Conservative: 64
Mismatches: 200
Indels: 112
Gaps: 19

US-10-014-896-1 (1-1509) x US-09-815-242-5375 (1-449)

```
Qy  163 CTGAAAGGTGCCATCCAGGATCCCAACAGTGACTTTTAGCTCTGAGAAGTCCAAAACTACA 222
        ||||||||||||   :::           :::|||   |||  :::  :::
Db    1 LeuLysGlyLeuLeuAlaIleGluSerValArgAspAspAlaSerGluAsp---        19

Qy  223 GCCCTGGGCTGCGAGTTCGGA------------------AAATACATTCATAAAGTCTTTCCTACAGTG 270
        ||||||   :::   |||           |||
Db   20 -----AlaProValGlyProGlyProArgGlyLysAlaLeuGluAspTyrMetTyrGlnIleLeuAla  37

Qy  271 GTCAGCCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAGTATAGCCACCTGGTCACTATC 330
        |||   ||| :::  :::                  :::
Db   38 HisArgAspGlyPheThrHisAlaGlyValAlaHisIleAlaGlyAlaGlyTyrIleGlyLeuGly    57

Qy  331 CAAGGCTGCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCCTCACTTTGAGTGTGGCCT 390
        |||   :::|||   :::                     |||:::
Db   58 LysGlyLysAspAsp------------------ValLeuGlyIleLeuCysHisSerValAspSerValValPro  73

Qy  391 GCCCCTGAACAAAGGCTGGGAGGGTCGCCCCCANTTCTCTGGGTTGGAGCGGTGATGGGGTCATC 450
        |||||||  :::                  ||| |||
Db   74 Ala----GlyAspGlyTyrProAspSerAsnProPheGluProValValThrGluAspAlaIle    92

Qy  451 TATGGTCGGGGCCACACTGGACGAACAAGAACTCTGTGAGTGGCATTACTGCAGGCCTGGAG 510
        |||||||||||||| :::||||||
Db   93 IleAlaArgGlyThrLeuGluAspProLysGlyGlyProThrIleAlaAlaLeuAlaTyrArgAlaLeuLys 112

Qy  511 CTCCTGCCTGATCAGGAAGTACATCCCCGAAGAGATCTTTCTTCCATTTCTGGGCCCATGAT 570
        :::|||                   |||   ||||||:::
Db  113 IleLeuGluAspMetAlaAsnValAspTrpLysArgGlnIleHisSerMetIleIleGlyThrAsp 132

Qy  571 GAGGAGTCATCAGGGACAGGACAGGGGCTCAGAG---------ATCTCAGCGCCTG 612
        |||||||  :::||||
Db  133 GluGluSerAspTrpLysCysThrAspArgGlyThrPheLysThrGlyLeuMetProThrLeu 152

Qy  613 CTACAGTCAAGGGGGCGGGGCGTCCAACTAGCCTTCATTGTGGACGAGGGGGGTCATC 672
        |||                     :::|||
Db  153 GlyPheAlaProAspAlaGluPheArgCysIleHisIsGlyGluLysLysGlyIleThrThrPhe 172

Qy  673 GATTTCATTCCTAAC----------------------- 687
        |||   :::   |||
Db  173 AspLeuValGlnAsnLysLeuThrThrGluGlnAspGluProAspTyrGluLeuThr 192

Qy  688 TTCAAGAAGCCCATCCGCCTGATCAGAGAAGGTTCCATGAACCTCATGCTG 747
        ||||||   :::
Db  193 PheLysSerGlyGluArgTyrAsnMetValProAspHisAlaGluAlaArgValLeuVal 212

Qy  748 CAAGTAAACATGACT----------------------------------------- 762
```

```
Db  170 ValValSerThrGlyProThrAspCysGlnValTyrArgGlyLysArgMetGlu 189
        :::  :::                              |||
Qy  739 CTCATGCTGCAAGTAAAACATGACTTCAGCCACTCTTCAGCTCCCTCCAAAGGAGACAAGC 798

Db  190 IleArgIleAspValGlnGlyValSerCysHisGlySerAlaProGluArgGlyAspAsn 209
        |||         :::          |||
Qy  799 ATTGGCATCCTTGCAGCTCGTCTGTGCAGCCATTGAGCAGCAGACACCAATGCCTATCATTT 858

Db  210 --------------------------------------------AlaIlePhe 212
        |||||||
Qy  859 GGAAGCGGGACAGTGGTGGTGATGTGTATTGCAGCAACTGGCAAATGAGTTTCCCTTCCCTGTC 918

Db  213 LysMetGlyProIleLeuGlyLeuGlnGluLeuSerGlnArgLeuGlyTyr----- 230
        :::
Qy  919 AATATAATCTGAGCAACCCATGGCTATTGTGAACCACTTATAAGCAGGTTTATGGAGAGA 978

Db  231 -----------------------------------------AspGluPheLeuGlyLys 236
        |||:::
Qy  979 AATCCCTTAACCAATGCAATAATCAGGAGACCACCACGGCCACTCACCATATTCAAAGCAGGG 1038

Db  237 Gly-------------------ThrLeuThrValSerGluIlePhePheThrSer 248
        |||   |||
Qy 1039 GTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGCCCACAGTCAACTTCCAGGATTCACCCT 1098

Db  249 ProSerArgCysAlaLeuAlaLeuAspSerCysAlaValSerIleAspArgArgLeuThrTrp 268
        |||:::
Qy 1099 GGACAGACAGTCCAAGAGGTTCCTAGAACTCACGAAGAACATTGTGGCTGATAACAGAGTC 1158

Db  269 GlyGluThrTrpGlu------ 273
        ::::::
Qy 1159 CAGTTCCATGTGTTGAGTGCCTTTGACCCCTCCCGGTCAGCCCTTCTGATGACAAGGCC 1218

Db  274 --------------GlyAlaLeuAspGluIleArgAlaLeuProAlaValGlnLysAla 288
        :::   |||||||
Qy 1219 TTGGGCTACCAGCGCTCTCCGCACCGTACAGTCCGTCTTCCCGGAAGTCAAATATTACT 1278

Db  288 ------------------------------------------- 288
Qy 1279 GCCCCAGTTACTTCTATTGGCAACACAGCAGCCGATTCTTTACAAAACTTCACCACTGGC 1338

Db  289 AsnAlaValAlaSerMetTyrAsnThrAspArgProSerTrpThrGlyLeuValTyrPro 308
        |||         |||
Qy 1339 ATCTACAGGTTCTACCCCATCTACACACAGCCTGAAGACTTCAAACGACCATGGAGTC 1398

Db  309 ThrGluCysTyrTyrPheProThrThrpLysValGluAspHisPheThrVal------ 325
        :::::::||||||   :::
Qy 1399 AACAGAGAAAAATCTCAGTCCAAAGCCTATGAG 1428

Db  326 ------LysAlaLeuValAsnAlaTyrGlu 333
        |||||||
```

RESULT 9
US-09-815-242-12290
; Sequence 12290, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23

; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SEQ ID NO: 12290
; SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 12290
LENGTH: 469
TYPE: PRT
ORGANISM: Staphylococcus aureus
US-09-815-242-12290

Alignment Scores:
Pred. No.:
Score: 0.00027
Percent Similarity: 145.00
Best Local Similarity: 35.00%
Query Match: 21.67%
DB: 5.25%
                          10

Length: 469
Matches: 104
Conservative: 64
Mismatches: 200
Indels: 112
Gaps: 19

US-10-014-896-1 (1-1509) x US-09-815-242-12290 (1-469)

```
Qy   163 CTGAAAGGTGCCATCCAGATTCCAACAGTGACTTTAGCTGTGAGAAGTGCAATACTACA 222
             |||||||  ::: |||  :: || ||::  |||
Db    17 LeuLysGlyLeuLeuAlaIleGlnSerValArgSerProAlaSerGluAsp---     35

Db   223 GCCCTGGCTGAGTTCGGA----------AAATACATTCATAAAGTCTTCCTACAGTG 270
                                  0
Qy    36 -----AlaProValGlyProGlyProArgLysAlaLeuAspTyrMetTyrGluIleAla  53

Qy   271 GTCAGGACCAGCTTTATCCACGATGAAGTCCGTGGAAGAGTATAGCCACCGTTCACTATC 330
Db    54 HisArgAspGlyPheThrThrHisAspValHisIleAlaGlyArgIleGluAlaGly     73

Db   331 CAAGGCTCGGACCCCAGCTTGCAGCCCTACCTTGCTGATGGCTCACTTTGAGTGTGGCCT 390
            |||  ||||:::|||      |||:: |||      |||::||:  ||
Qy    74 LysGlyAsnAsp--------------ValLeuGlyIleLeuLeuCysHisSerValAspVal 89

Qy   391 GCCCCTGAAAGAAGGTGGGAGGGTGCCCCAATTCTCTGGGTTGGAGCGTGGCGTCATC 450
         ||| |||         ||:: |||:: ||        ||| ||
Db    90 Ala-----GlyAspGlyTyrProAlaPheSerAsnProPheGluProValValThrGluLeu 108

Db   451 TATGGTTCGGGCGCACTGGACGACAACGAAGAACTTCTGTGATGGCCATTACTGACGGGTGGAG 510
Qy   109 IleAlaArgGlyThrLeuAspAspLysGlyProThrIleAlaAlaLaTyrArgTyrArgAlaLys 128

Qy   511 CTCCTTGCTGATCAGGAAGTACATCCCCGAAGAGATCTTTCTTCATTTCTCTGGGGCCATGAT 570
         :::::     |||:: ::  :::|| ::::
Db   129 IleLeuGluAspMetAsnValAspTrpLysGlyLysArgIleHisMetIleIleGlyThrAsp 148

Qy   571 GAGGAGTCATCAGGGACAGCAGGGGCTCAGAGG--------------ATCTCAGCCCTG 612
         |||||||
Db   149 GluGluSerAspTrpLysCysThrAspArgTyrPheLysThrArgLeuGluMetProThrLeu 168

Db   613 CTACAGTCAAGGGGCGGGCTCCAGCTACGCCTTCAATTGTGGGACGAGGGGGGGTCATC 672
             :::                   |||                        |||
Qy   169 GlyPheAlaProAspAlaGluProCysIleHisGlyGluTyrGlyIleThrHisThrPhe 188

Qy   673 GATTTCATTCCTAAC--------------------------------- 687
         |||         |||
Db   189 AspLeuValGlnAsnLysLeuThrGluThrGlnIleAsnGlnProAspTyrGluLeuIleThr 208

Db   688 TTCAAGAAGCCCATCCGCCTTGATTGCGACGTCTCAGAAGTGCAGAGAGGGTCCCATGAACCTTCATCATG 747
            ||||||  |||
Qy   209 PheLysSerGlyLeuArgTyrAsnMetValProAspHisAlaGluAlaArgValLeuVal 228

Qy   748 CAAGTAAAACAATGACT--------------------------------- 762
```

```
Db    229  :::||||||| |
           LysGluAsnMetThrAspValIleGlnAspPheGlyThrTyrPheLeuGluGlnAsnHisLeu 248
Qy    763  TCAGGCCACTCTTCAGCTCCTCCAAAGGAGACAAGCATTGGCATCCTTGCAGTCGCTGTC 822
           |||   ||:::
Qy    249  GlnGlyAspSerThr------ValAlaSpSerGlyIleLeuValLeuThrVal 263
           |||||||| |||
Db    823  AGCCGATTGGAGCAGAGACCAATGCCTATCATATATTGGAAGCGGACAGTGGTGACTGTA 882
Db    264  GluGlyLysAlaValHisGlyMetAspProSerIleGlyValAlaSnAlaGlyLeuTyrLeu 283
           :::
Qy    883  TTGCAGCAACTGGCA------AATGAGTTTCCCTTGCCTTGCTCAATATA 924
           |||||:::       :::
Qy    284  LeuLysPheLeuAlaSerLeuAsnLeuAspAsnAsnAlaGlnAsnAlaPhe------Val 300
Db    925  ATCCTGAGCAACCCATGGCTATTTGAACCACTTATAAGCAGGTTTATGGAGAGAAAATCCC 984
           |||||||| |||    :::
Db    301  AlaPheSerAsnArgTyrLeuPheAsn------SerAspPheGlyGlu--- 314
Qy   1045  TTC-------AATGTCATCCCCCCAGTGGCCCAGGCCACA------ 1077
           |||         :::
Qy    315  ------LysMetGlyMetLys 319
Db    985  TTAACCAATGCCAATAATCAGGACCACCACGGCACTCACCATATTCAAAGCAGGGGTCAAG 1044
Db    320  PheHisThrAspValMetGlyAspValThrThrAsnIleGlyLValIleThrTyrAspAsn 339
Qy   1078  -------GTCAACTTCCGGATTCACCCTGGACAGAGACAGTCCAAGAG 1116
           :::|||||||
Db    340  GluAsnAlaGlyLeuPheGlyIleLeuAsnLeuArgTyrProGluGlyLypheGluGluLys 359
Qy   1117  GTCCTAGAACTCACGAAGAACAATTGTGGCTGAAATAACAGAGTCCAGTCACATGTGTTGAGT 1176
           :::  |||          :::
Db    360  AlaMetAspArgPheLaAsnGlnIleLeuGlnTyrGlyPheGluMetLysLeuGlyLys 379
Qy   1177  GCCTTTGACCCCCTCCCCGTCAGCCCTTCTGATGACAAGGCCTTGGGCTACCAGCTGGCTC 1236
           |||:::       :::
Db    380  ValGlnProProHisThrTyrValAlaSpLysAsnAsp------ProPheVal 393
Qy   1237  CGCCAGACCGTACAGTCCGTCGTTCCGGAGTCAAT---AATTACTGCCCCCAGTTACTTCT 1293
           :::|||           :::
Db    394  GlnLysLeuValThrAlaTyrArgAsnGlnThrAsnAsPMetThrGluProTyrProThrIle 413
Qy   1294  ATTGGCAACACAGACAGCAGCCGATTCTTTACAAAACCTCACCACTGGCATCACAGGTTCTAC 1353
           |||  |||          :::
Db    414  GlyLyGlyLyThrTyrAlaArg------AsnLeuAspLysGlyVal---AlaPheGly 429
Qy   1354  CCCATCTACAATACAGCCTGAAGACTTCAAAACGACTGTCAACAGAGAAAAATCTCA 1413
           :::|||          :::||||||  :::|||
Db    430  AlaMetPheSerAspSerGluAspLeu------MetHisGlyLnLysAsnGluTyrIleThr 447
```

Alignment Scores:
Pred. No.:   0.00157   Length: 457
Score: 136.50   Matches: 80
Percent Similarity: 37.57%   Conservative: 50
Best Local Similarity: 23.12%   Mismatches: 155
Query Match: 4.94%   Indels: 61
                                            Gaps: 15

US-10-014-896-1 (1-1509) x US-09-738-626-6496 (1-457)

```
Qy    190  GTGACTTTTAGCTCTCGAGAAGTCCAATACTACAGCCCTGGCTGAGTCGGAAAATACATT 249
           |||:::|||::::|||                :::|||    :::
Db     29  ValSerPheAsnSerValHisSerAspSerProAsnLeuLeuGluAspTyrAlaGlyAlaLys 48
Qy    250  CATAAAGTCTTTCCTACAGTGGTCACCACCAGCTTTATCCAGCATGAAGTCGTGGAAGAG 309
           |||:::       :::
Db     49  GluTrpValLysGluThrThrLeuThrAsnAlaGlyLeuThrValSerGluPheAlaAlaGlu 68
Qy    310  ------TATAGGCACCNTGTTCACTATCCAAGGCTCGGACCCCAGCTTGCAGCCCTAC 360
           :::       |||
Db     69  AspGlyThrThrAsnPheIleGlyThrArgGlySerGlu---GlyAlaProLysVal 87
Qy    361  CTGCTCGATGGCTCACTTTGATGTGGTGCCGCCTGGCC---CCTGAAGAAGGCTGGGAGGTGCCC 417
           :::|||||||::::                 :::
Db     88  LeuLeuTyrSerHisPheAspValProSerGlyProLeuAspLeuTrpAspThrAsn 107
Qy    418  CCAATTCTCTGGGGTTGGGACGGTGATGGC------GTCATCTATGGTGCGGGGCACA 465
           |||            :::|||
Db    108  ProPheGluLeuThrGluArgAspAlaLeuGlyThrHisGlyThrArgTrpTyrGlyArgGlyAla 127
Qy    466  CTGGACGGACAAGAACTCTGTGATGGCAATTACTGCAGGTTG------ 507
           |||::::::|||            :::
Db    128  AlaAspCysLysGlyAsnLeuValMetHisLeuArgAlaValGluAlaSer 147
Qy    508  ---GAGCTCCTGCTGATCAGGAAGTACAATCCCCGAAGATCTTCTTCATTTCTGTGGGC 564
           :::|||          :::
Db    148  GlyAspThrThrLeuAsnLeuThrTyrValValGluGlySer------ 161
Qy    565  CATGATGAGGAGTCATCAGGGACAGGGGGCTCGAGACCCTGCTACAGTCAAGG 624
           :::||||||:::::|||
Db    162  ------GluGluMetGlyGlyGlyAla------LeuSerAlaLeuIleLysAspLys 176
Qy    625  ------GGGGTCCAGCTAGCCTTCATTGTGGACAGGGGGCTTCATCTTGGAT 672
           |||               :::
Db    177  ProGluLeuPheAspAlaAspSerValIleLeuIleLeuAlaAspSerGly--- 191
Qy    673  GATTTCATTCCTAACTTCAAGAAGCCCATCGCCTTGAATGCAGTCTCAGAGAAGGGTTCC 732
           |||      :::|||
Db    192  ------AsnAlaSerValGlyThrProThrLeuThrThrLeuArgGlyGly 207
Qy    733  ATGAACCTCAATGCTGCAAGTAAACATGACTTCAGGC------CACTCT--- 774
           |||:::    |||                 :::
Db    208  GlyInValThrValThrValAspThrLeuGluGlyValAlaValHisSerGlyGlnTyrGly 227
Qy    775  ---TCAGCTCCTCCAAAGGAGACAAGCATTGGCATCCTTGCAGTCGCTGTC 831
           :::|||||||
Db    228  GlyAlaIlaAlaPro------AspAlaIlaAlaLeuValArgValLeuAspThrLeu 244
Qy    832  GAGCAGCACCAAATGCCTATCATATATTGGAAGCGGGACAGTGGTGACTGTATTGCAGCAA 891
           |||||||:::||||||                :::
Db    245  ArgAspGlu------HisGlyArgThrValIleAspGlyValAsnThr 258
```

```
Qy    892  CTGGCAAAT------GAGTTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTA  945
           ||||||                  ||||:::
Db    259  ThrAlaAsnTrpLysGlyGluProTyrAspProGluThrPheArgSerAspAlaGlyIle  278

Qy    946  TTTGAACCACTTATAAGCAGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGG  1005
           |||||||                                   :::
Db    279  LeuAspGlyValAspIleMetGlyAspAlaAspAsnProAlaSerMetLeuTrpSerArg  298

Qy    1006 ACCACCACGGCACTCACCATATTCAAAGCA------GGGGTCAAGTTCAAATGTCATC  1056
           |||||||||||||||||||||||                       :::
Db    299  ProAlaIleSerIleThrGlyPheThrSerThrProValAlaGlyAlaAlaLeuAsnAlaVal  318

Qy    1057 CCCCCAGTGGCCCCAGGCCCAGTCAACTTCCGGATCACCCTGGACAGACAGTCCAAGAG  1116
           |||                     |||          ::::
Db    319  ProAlaThrAlaSerAlaLysLeuAsnLeuArgValProAlaGlyLeuGluAlaAsnAsp  338

Qy    1117 GTCCTAGAACTCACGAAG  1134
           |||   |||    |||
Db    339  ValAlaGlyLysLeuLys  344
```

RESULT 11
US-09-791-389-2
; Sequence 2, Application US/09791389
; Publication No. US20030032773A1
; GENERAL INFORMATION:
;   APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri
;   APPLICANT: Parekh, Rajesh Bhikhu
;   APPLICANT: Rohlff, Christian
;   APPLICANT: Terrett, Jonathan Alexander
;   APPLICANT: Tyson, Kerry Louise
;   TITLE OF INVENTION: Proteins, Genes and Their Use for
;   TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)
;   TITLE OF INVENTION: and Unipolar Depression
;   FILE REFERENCE: 2543-1-001 N2
;   CURRENT APPLICATION NUMBER: US/09/791,389
;   CURRENT FILING DATE: 2001-02-23
;   PRIOR APPLICATION NUMBER: GB 0004412.3
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: GB 0030050.9
;   PRIOR FILING DATE: 2000-12-08
;   PRIOR APPLICATION NUMBER: US 60/254,830
;   PRIOR FILING DATE: 2000-12-12
;   NUMBER OF SEQ ID NOS: 308
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 501
;   TYPE: PRT
;   ORGANISM: homo sapien
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(501)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-09-791-389-2

Alignment Scores:
Pred. No.:   0.00164
Score: 136.50
Percent Similarity: 37.30%
Best Local Similarity: 21.62%
Query Match: 4.94%
DB: 9

                          Length: 501
                          Matches: 80
                          Conservative: 58
                          Mismatches: 127
                          Indels: 105
                          Gaps: 17

US-10-014-896-1 (1-1509) x US-09-791-389-2 (1-501)

Qy   334 GGCTCGGACCCCAGCTTGCCAGCCCTACCTGCTGAATGGCTCACTTTGATGTGGTGCCCTGCC 393
           |||||||||||||||||||||||||||||||||||::::::: ||
Db   113 GlySerAspProThrLysGlyThrValCysPheTyrGlyHisLeuAspValGlnProAla 132

Qy   394 CCTGAA---GAAGGCGTGGGAGAGTGCCCCCATTCTCTCGGGTTGGAGGCGTGAGGGCGTCATC 450
           :::|||||| ||||||
Db   133 AspArgGlyLysAspGlyTrpLeuThrArgAspProTyrValLeuThrHisGluValAspGlyLysLeu 152

Qy   451 TATGGTCGGCGGGCACACTGGACGACAAGAAACTCTGTGATGGCATTACTACTGCAGGCCTGGAG 510
```

```
           ||||||||||   ||:::::||
Db    153  TyrGlyTyrArgGlyValAlaThrArgAspAsnLysGlyGlyProValLeuAlaTrpIleAsnAlaValSer  172

Qy    511  CTCCTG---CTGATCAGGAAGGTACATCCCCGGAAGATCTTTCTTACTTGGCCAT  567
           |||:::        :::                      |||||||    ||||||
Db    173  AlaPheArgAlaLeuGluGlnAspLeuProValAsnIleLysPheIleLeuGluGlyMet  192

Qy    568  GATGAGGAGTCATCAGGGACAGGGGCTCAGGAGGAGTCAGGCCCGCTACAGTCAAGG---  624
           :::||||  |||                |||:::||          ||||:::||
Db    193  GluGlu---------AlaGlySerValAlaLeuGluGluValGluLysGlyLys  208

Qy    625  -----GGGCTCCAGCTAGCCTTCATTGTGGACGAGGGGGT  669
                |||||||              ||||:::
Db    209  AspArgPheSerGlyValAsp-----TyrIleVal-----  221

Qy    670  GATGAGTTCATTCCTAACTTCAAGAAGCCCATCGCCTTGATTGCAGTCTCAGAGAAGGGT  729
           |||:::                |||              ||||||
Db    222  AspAsnLeuTrpIleSerGlnArgLysPro-----AlaIleThrTyrGlyThrArgGly  239

Qy    730  TCCATGAACCTCATGCTGCAAGTAAACATGACTTCAGGC------CACTCTTCAGCTCCT  783
           |||:::                                  ||||||
Db    240  AsnSerTyrPheMetValGluValLysCysArgAspGlnAspPheHisSerGly-----  257

Qy    784  CCAAAGGAGACAAAGCATTGGCATCCTTGCAGCCTGCCTGTCTCAGCCGGATTGGAGCAGCACCA  843
           ||||||                                            |||
Db    258  -----ThrPheGlyIleTyrIleLeuHisGluProMetAlaAspLeuValAlaLeuAspLeu  274

Qy    844  ATGCCTATCATATTTGGAAGCGGGACAGTGGTGACTGTATTGCAGAACTGGCAAATGAG  903
           ||||||                                          |||:::
Db    275  GlySerLeuValAspSerSerGlyHisIleLeuVal-----  286

Qy    904  TTTCCCTTCCCTGTCAATATAATCCTGAGCAACCCATGGCTATTTGAACCACTTATAAGC  963
           ||||||                             ||||||:::
Db    287  ---------ProGlyIleTyrAspGluValVal-----  294

Qy    964  AGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATCAGGACCACCACGGCACTCACC  1023
           |||||||                            |||
Db    295  ----ProLeuThrGlyIleGluIleLeuAsnThr-----  303

Qy    1024 ATATTCAAAGCACAGGGGGTCAAGTTCAAGTTCAATCCCCCAGTGGCCCCAGGCCCACAGTCAAC  1083
           ||||||                               |||
Db    304  --TyrLysAla-----  306

Qy    1084 TTCCGGATTCAC------CCTGGACAGACAGTCCAAGAGGTCCTA  1122
           |||:::                |||||:::
Db    307  -----IleHisLeuAspLeuGluGluTyrArgAsnSerSerArgValGluLysPheLeu  324

Qy    1123 GAACTCACGAAGCAACATTGTGGCTGATAAACAGAGTCCCAGTTCCACGTGTTGGAGT---  1176
           |||||||                                          |||:::
Db    325  PheAspThrLysGluIleLeuMetHisLeuThrProArgTyrProSerLeuSerIleHis  344

Qy    1177 ----GCCTTTGACCCCTCCAGCCCTTCGATGACAAGGCCTTGGGC  1224
                ||||||||||
Db    345  GlyIleGluGlyValAlaPheAsp-----GluProGlyThrLysThrValIlePro  360

Qy    1225 TACCAGCTGCTCCGCGACGACTACAGTCCGTCTCCCGGAAGTCAATATTACTGCC---  1281
           ||||||:::                :::
Db    361  GlyArgValIleGlyLysPheSerIleArgLeuValProHisMetAsnValSerAlaVal  380

Qy    1282 ------CCAGTTACTTCTATTGGAACACAGAGACAGAGCCGA  1314
                :::
Db    381  GlyLysGlnValThrArgHisLeuGluGluAspValPheSerLysArgAsnSerSerAsnLys  400

Qy    1315 TTCTTTACAAACCTCACCACTGGCATCTAC  1344
           |||:::||:|| |||||:::
Db    401  MetValSerMetThrLeuGlyLeuHis  410
```

RESULT 12
US-09-791-393-2
; Sequence 2, Application US/09791393
; Publication No. US20030032200A1
; GENERAL INFORMATION:
;   APPLICANT: Herath, Herath Mudiyanselage Athula Chandrasiri

```
;  APPLICANT: Parekh, Rajesh Bhikhu
;  APPLICANT: Rohlff, Christian
;  TITLE OF INVENTION: Proteins, Genes and Their Use for
;  TITLE OF INVENTION: Diagnosis and Treatment of Bipolar Affective Disorder (BAD)
;  TITLE OF INVENTION: and Unipolar Depression
;  FILE REFERENCE: 2543-1-001 N1
;  CURRENT APPLICATION NUMBER: US/09/791,393
;  CURRENT FILING DATE: 2002-01-02
;  EARLIER APPLICATION NUMBER: GB 0004412.3
;  EARLIER FILING DATE: 2000-12-08
;  EARLIER APPLICATION NUMBER: GB 0030050.9
;  EARLIER FILING DATE: 2000-02-24
;  EARLIER APPLICATION NUMBER: GB 0030050.9
;  EARLIER FILING DATE: 2000-12-08
;  EARLIER APPLICATION NUMBER: US 60/254,830
;  EARLIER FILING DATE: 2000-12-12
;  NUMBER OF SEQ ID NOS: 308
;  SOFTWARE: FastSEQ for Windows Version 4.0
;
;  SEQ ID NO 2
;    LENGTH: 501
;    TYPE: PRT
;    ORGANISM: homo sapien
;    FEATURE:
;    NAME/KEY: VARIANT
;    LOCATION: (1)...(501)
;    OTHER INFORMATION: Xaa = Any Amino Acid
DB: US-09-791-393-2

US-10-014-896-1 (1-1509) x US-09-791-393-2 (1-501)

Alignment Scores:
Pred. No.:
Score:              0.00164          Length:       501
Percent Similarity: 136.50           Matches:      80
Best Local Similarity: 37.30%        Conservative: 58
Query Match:        21.62%           Mismatches:   127
                    4.94%            Indels:       105
                    9                Gaps:         17
```

```
Qy   334 GGCTCGGACCCCAGCTTGCAGCCCTACCTGCTGATGGCTCACTTTGATGTGGTGCCTGCC 393
         ||||||||||||||||||||:::|
Db   113 GlySerAspProThrLysGlyThrValCysPheTyrGlyHisLeuAspValGlnProAla 132

Qy   394 CCTGAA---GAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTGGAGCGTGATGCGTCATC 450
         ::|||||||                  ||:|||
Db   133 AspArgGlyAspGlyTrpLeuThrAspProTyrValLeuThrGluValAspGlyLysLeu 152

Qy   451 TATGGTCGGGGCCACTGGACGACAAGAAACTCTGTGATGGCATTACTGCAGGCCTTGGAG 510
         ||||||||||||               |||:::|||               |||:::
Db   153 TyrGlyArgGlyAlaThrAspAsnLysGlyProValLeuAlaTrpLeuAsnAlaValSer 172

Qy   511 CTCCTG---CTGATCAGGAAGTACATCCCCGGAAGATCTTCTTCAATTCTCTGGGCCAT 567
         |||||||            ::|               |||:::
Db   173 AlaPheArgAlaLeuGluGlnAspLeuProValAsnIleLysPheIleIleGluGlyMet 192

Qy   568 GATGAGGAGTCATCAGGGACAAGGGCCTCAGAGGATCTCAGCCCTGCCTACAGTCAAGG--- 624
         ||:|||||            :::||||||
Db   193 GluGlu--------          |||:::|||
                          AlaGlySerValAlaLeuGluGlnLeuValGluLysGluLys 208

Qy   625 --------          ::|||||||
         ||:|||||
Db   209 AspArgPhePheSerGlyValAsp----------TyrIleVal----------IleSer 221

Qy   670 GATGATTTCATTCCTAACTTCAAGAGAGCCCATCGCCTTGATTGCAGTCTTCAGAGAAGGGT 729
         |||:::||            :::|||||||
Db   222 AspAsnLeuTrpIleSerGlnAsnArgLysPro----------AlaIleThrTyrArgGly 239

Qy   730 TCCATCGAAACCTCATCGTGCAAGTAAACATGACTTCAGGC----CACTCTTCAGCTCCT 783
         :::|||||            :::||||||
Db   240 AsnSerTyrPheMetValGluValLysCysArgAspGlnAspPheHisSerGly------ 257

Qy   784 CCAAGGAGACAAAGCAATTGGCATCCTTGCAGCTGCTTGTCAGCCGATTGGACGACACCA 843
         |||:::              |||||||||
Db   258 --------          |||
         ThrPheGlyGlyIleLeuHisGluProMetAlaAspLeuValAlaLeuLeu 274

Qy   844 ATGCCTATCATCATATTTGGAAGGGGGACAGTGGTGACTGTATTGCAGCAACTGGCAAATGAG 903
```

```
Db   275 GlySerLeuValAspSerSerGlyHisIleIleLeuVal----------           286
         ||||||                   ::::::

Qy   904 TTTCCCTTCCCTGTCAATATAATNAATCCTGAGCAACCACCATGGCTATTTGAACCACTTATAAGC 963
                                          ||
Db   287 ----------                                               294
         ----ProGlyIleTyrAspGluValVal----

Qy   964 AGGTTTATTGGAGAAAAATCCCTTAACCAAATGCAATAATCAGGACCACCACGGCCACTCACC 1023
         ||||||||||                                  ||||||
Db   295 ----------          |||                         294
         ProLeuThrGluGlnIleIleAsnThr----                    303

Qy  1024 ATATTCAAAGCAGGGGGTCAAGTTCAAATGTCATCATCCCCCAGTGCCCAGGCCACAGTGCAAC 1083
         ||||||||||||                          |||
Db   304 ----------          |||                           306
         ----TyrLysAla----

Qy  1084 TTCCGGATTCAC---------- 1122
         ||||||
Db   307 ----------          ||||||         324
         ----IleHisLeuAspLeuGluGlnThrTyrArgAsnSerSerArgValGluLysPheLeu

Qy  1123 GAACTCACGAAGAACAATTGTGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGT---- 1176
         :::||||||         :::|
Db   325 PheAspThrLysGluGlnIleIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis 344

Qy  1177 ----------          ||||||         1224
         ----GCCTTTGACCCCCTCCCGGTCAGCCCCTTCTGATGACAAGGCGTCCTTGGGC
Db   345 GlyIleGluGlyIleGlyAlaPheAsp----------           360
         ----GluProGlyThrLysThrValIlePro

Qy  1225 TACCAGCTGCTCCGCCAGACCGTACAGTCCGCTGTCTTCCGGGAAGTCAATAATTATTACTGCC--- 1281
         :::::::|||         :::::::|
Db   361 GlyArgValIleGlyLysPheSerIleArgLeuValProHisMetAsnValSerAlaLeuVal 380

Qy  1282 ----------          |||         1314
         ----CCAGTTACTTCTTCTTATTGGCAACACAGAACAGCCGA
Db   381 GluLysGlnValThrArgHisLeuGluGlnPheSerLysArgAsnSerSerAsnLys 400

Qy  1315 TTCTTTACAAACCTCACCACCTGGCCATCTAC 1344
         |||:::::::
Db   401 MetValValSerMetThrThrLeuGlyLeuHis 410
```

```
RESULT 13
US-09-791-378-674
;  Sequence 674, Application US/09791378
;  Patent No. US20020142303A1
;  GENERAL INFORMATION:
;  APPLICANT: Parekh, Rajesh
;  TITLE OF INVENTION: PROTEINS, GENES AND THEIR USE FOR DIAGNOSIS AND TREATMENT OF
;  TITLE OF INVENTION: SCHIZOPHRENIA
;  FILE REFERENCE: 9195-061-999
;  CURRENT APPLICATION NUMBER: US/09/791,378
;  CURRENT FILING DATE: 2001-02-23
;  PRIOR APPLICATION NUMBER: 09/750,395
;  PRIOR FILING DATE: 2000-12-28
;  NUMBER OF SEQ ID NOS: 677
;  SOFTWARE: Patentin version 3.0
;  SEQ ID NO 674
;    LENGTH: 501
;    TYPE: PRT
;    ORGANISM: Homo sapiens
;    FEATURE:
;    NAME/KEY: MOD_RES
;    LOCATION: (70)...(70)
;    OTHER INFORMATION: Xaa = Ile or Leu
DB: US-09-791-378-674

US-10-014-896-1 (1-1509) x US-09-791-378-674 (1-501)

Alignment Scores:
Pred. No.:
Score:                 0.00164        Length:       501
Percent Similarity:    136.50         Matches:      80
Best Local Similarity: 37.30%         Conservative: 58
Query Match:           21.62%         Mismatches:   127
                       4.94%          Indels:       105
                       10             Gaps:         17
```

```
Qy    334  GGCTCGGACCCCAGGTTGCAGCCCTACCTGCTCGATGGGCTCACTTGATGGGTGCCTGCC  393
           ||||||||||||||||||||:::
Db    113  GGTGGCGAGGTGGGCCCCCATTCTCTGGGTTGGAGCGGTGATGGGTCATC  132

Qy    394  CCTGAA----GAAGGCTGGGAGGGTGCCCCATTCTCTGGGTTGGAGCGGTGATGGGTCATC  450
           |||:::                  |||:|
Db    133  AspArgGlyAspGlyTrpLeuThrAspProTyrValLeuThrGluValAspGlyLysLeu  152

Qy    451  TATGGTCGGGGCCCGGCACTGGACGACAAAGAACTCTGTGATGGCATTAGTACTGGGAG  510
           ||||||||||||          ::|||||         |||::|
Db    153  TyrGlyArgGlyAlaThrAspAsnLysGlyProValLeuAlaValSer  172

Qy    511  CTCCTG----CTGATCAGGAAGTACAATCCCCGAGATCTTTTTCTTCATTTCTTCTGGGCCAT  567
           ||||||                 :::||                    |||||
Db    173  AlaPheArgAlaLeuGluGlnAsnAspLeuProValAsnIleLysPheIleIleGluGlyMet  192

Qy    568  GATGAGGAGTCATCAGGAGACAGGGGCTCAGGAGGATCTCAGCCCTGCTACAGTGTCAAGG---  624
           ::|||                 |||
Db    193  GluGlu------AlaGlySerValAlaLeuGluLeuValGluValGlyLysLys  208

Qy    625  -------GGCGGTCCAGCTTAGCCCTTCATTGTGGACGAGGGGGGGCCTTCATCTTG  669
                  |||||               ::|||||||
Db    209  AspArgPhePheSerGlyValAsp------TyrIleVal  221

Qy    670  GATGGATTTCATTCCTAACTTCAAGAGAGCCCATACGGCCTTTGATTGATGCAGTCTCAGAGAGGGGT  729
           |||:::                ::|:||||||                ::||||||
Db    222  AspAsnLeuTrpIleSerGlnArgLysPro------AlaIleThrThrArgGlyThrArgGly  239

Qy    730  TCCATGAAACCTTCATGCTGTGCAAGTAAAAACATGACTTCAGGC------CACTCTTCAGCTCCT  783
           :::                        :::                |||||
Db    240  AsnSerTyrPheMetValGluValLysCysArgGlyAspAspHisSerGly------  257

Qy    784  CCAAAGGACAGACAAGCAATGGGACATCCTTGCTGCTGTCTCCAGCCGATTGGAGCAGAGACACCA  843
           |||||                 :::|||||||
Db    258  ------ThrPheGlyIleLeuHisGluProMetAlaAspLeuValAlaLeuLeuLeu  274

Qy    844  ATGCCTATCATATTTGGAAGGCGGACAGTGGCTGGACTGTATTGCAGCAACTGGCAAATGAG  903
           |||:::                ::||||||                 ::||||||
Db    275  GlySerLeuValAspSerSerGlyHisIleLeuVal------AlaIleThrArgGlyThrArgGly  286

Qy    904  TTTCCCTTCCCTGTCAATATAATCTCTGAGCAAACCCATGGGTATTTGAAACCACTTATAAGC  963
           ::::||||||              :::|||
Db    287  ------ProGlyIleTyrAspGluValVal---  294

Qy    964  AGGTTTATGGAGGAGAAATCCCTTAACCAATCCAAATAATCAGGACCACCACGGCCACTCACC  1023
           ||||||||||:::                 |||
Db    295  ProLeuThrGluGluValIleLeuAsnThr------  303

Qy    1024 ATATTCAAAAGCAGGGGGTCAAGTTCAAATGTCATCCCCCAGTGGCCCAGCCACAGTCAAC  1083
           ::||||||                 :::|||
Db    304  ------TyrLysAla---  306

Qy    1084 TTCCGGATTCAC------CCTGGACAGAACAGTCCAAGAGGTCCTA  1122
           ||||||               |||
Db    307  ------IleHisLeuAspLeuGluGluTyrArgAsnSerArgValGluLysPheLeu  324

Qy    1123 GAACTCACGAGAGAACATTGTGGCTGATTAACAGAGTCCAGTGTTCCATGTGTTCGAGT----  1176
           ||||||               :::
Db    325  PheAspThrLysGluGluIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis  344

Qy    1177 ------GCCTTTGACCCCCTCCCCGCCGTCAGCCGTCTGATGACAAGGCCTGGGC  1224
           ::||||||                 :::
Db    345  GlyIleGluGlyValAlaPheAsp------GluProGlyThrLysThrValIlePro  360

Qy    1225 TACCAGCTGCTCCGCGACACCGTACAGTCCGTCTCTCCGGAGTCAATATTACTGCC---  1281
           ||||||               :::||||||
Db    361  GlyArgValIleGluGlyLysPheSerIleArgLeuValProHisMetAsnValSerAlaVal  380

Qy    1282 ------CCAGTTACTTCTTATTGGCAACACAGCAGCCGA  1314
           :::||||||              :::
Db    381  GluLysGlnValThrArgHisLeuLeuGluAspValPheSerLysArgAsnSerSerAsnLys  400
```

```
Qy    1315 TTCTTTACAAAACCTCACCACTGGGCATCTAC  1344
           :::::::||||                 ||:::::
Db    401  MetValValSerMetThrLeuGlyLeuHis  410
```

RESULT 14
US-10-036-041-57
; Sequence 57, Application US/10036041
; Publication No. US20020192751A1
; GENERAL INFORMATION:
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3030R1C8
; CURRENT APPLICATION NUMBER: US/10/036,041
; CURRENT FILING DATE: 2001-12-26
; PRIOR APPLICATION NUMBER: 60/085579
; PRIOR FILING DATE: 1998-05-15
; PRIOR APPLICATION NUMBER: 60/112514
; PRIOR FILING DATE: 1998-12-15
; PRIOR APPLICATION NUMBER: 60/113300
; PRIOR FILING DATE: 1998-12-22
; PRIOR APPLICATION NUMBER: 60/113430
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113605
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/113621
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/114140
; PRIOR FILING DATE: 1998-12-23
; PRIOR APPLICATION NUMBER: 60/115552
; PRIOR FILING DATE: 1999-01-12
; PRIOR APPLICATION NUMBER: 60/116843
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/125774
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125778
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: 60/125826
; PRIOR FILING DATE: 1999-03-24
; PRIOR APPLICATION NUMBER: 60/127035
; PRIOR FILING DATE: 1999-03-31
; PRIOR APPLICATION NUMBER: 60/127706
; PRIOR FILING DATE: 1999-04-05
; PRIOR APPLICATION NUMBER: 60/129122
; PRIOR FILING DATE: 1999-04-13
; PRIOR APPLICATION NUMBER: 60/130359
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/131270
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131272
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/131291
; PRIOR FILING DATE: 1999-04-27
; PRIOR APPLICATION NUMBER: 60/132371
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132379
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/132383
; PRIOR FILING DATE: 1999-05-04
; PRIOR APPLICATION NUMBER: 60/135750
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/138166
; PRIOR FILING DATE: 1999-06-08
; PRIOR APPLICATION NUMBER: 60/144791

```
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: 60/146970
; PRIOR FILING DATE: 1999-08-03
; PRIOR APPLICATION NUMBER: 60/162506
; PRIOR FILING DATE: 1999-10-29
; PRIOR APPLICATION NUMBER: 09/311832
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/380142
; PRIOR FILING DATE: 1999-08-25
; PRIOR APPLICATION NUMBER: 09/644848
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 09/747259
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 09/816744
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/854208
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/854280
; PRIOR FILING DATE: 2001-05-10
; PRIOR APPLICATION NUMBER: 09/874503
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/869599
; PRIOR FILING DATE: 2001-06-05
; PRIOR APPLICATION NUMBER: 09/908,827
; PRIOR FILING DATE: 2001-07-18
; PRIOR APPLICATION NUMBER: PCT/US99/10733
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: PCT/US99/28551
; PRIOR FILING DATE: 1999-12-02
; PRIOR APPLICATION NUMBER: PCT/US99/30720
; PRIOR FILING DATE: 1999-12-22
; PRIOR APPLICATION NUMBER: PCT/US00/05601
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: PCT/US00/23522
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: PCT/US00/23328
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: PCT/US00/05841
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: PCT/US00/14042
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: PCT/US00/15264
; PRIOR FILING DATE: 2000-06-02
; PRIOR APPLICATION NUMBER: PCT/US00/32678
; PRIOR FILING DATE: 2000-12-01
; PRIOR APPLICATION NUMBER: PCT/US00/34956
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: PCT/US01/06520
; PRIOR FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: PCT/US01/17800
; PRIOR FILING DATE: 2001-06-01
; PRIOR APPLICATION NUMBER: PCT/US01/19692
; PRIOR FILING DATE: 2001-06-20
; PRIOR APPLICATION NUMBER: PCT/US01/21066
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: PCT/US01/21735
; PRIOR FILING DATE: 2001-07-09
; NUMBER OF SEQ ID NOS: 80
; SEQ ID NO 57
; LENGTH: 507
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-036-041-57
US-10-014-896-1

Alignment Scores:
Pred. No.:                   0.00165
Score:                       136.50
Percent Similarity:          37.30%        Matches:        80
Best Local Similarity:       21.62%        Conservative:   58
Query Match:                 4.94%         Mismatches:     127
DB:                                        Indels:         105
                             Length: 507   Gaps:           17

US-10-014-896-1 (1-1509) x US-10-036-041-57 (1-507)
```

```
Qy    334 GGCTCGGACCCCAGCTTGCAGCCCTTACCCTGCTGATGGGCTCCACTTTGATGTGGTCGCC 393
              |||||||||||||||:::     ::||   ||  |||   ||||||||||:::  |||||||
Db    119 GlySerAspProThrLysGlyThrValCysPheTyrGlyHisLeuLeuAspValGlnProAla 138

Qy    394 CCTGAA---GAAGGCTGGGAGGTGCCCCCATTCTCTGGGTTGGAGCGTGATGGGGTCATC 450
              :::|||||||
Db    139 AspArgGlyAspGlyTrpLeuThrAspProTyrValLeuThrLeuThrLeuValAspGlyLysLeu 158

Qy    451 TATGGTCGGGGGGCCACAGGCGACGAAGAACTTCTGTGATGGCCATTACTGGAG 510
              |||||||||
Db    159 TyrGlyArgGlyAlaThrAspAspGluGluLeuLeuTrpIleAsnAlaValSer 178

Qy    511 CTCCTG---CTGATCAGGAGTACATCCCCCGAGAGATCTTTTCTCATTTGTCTGGGCCAT 567
              :::     :::|||     |||
Db    179 AlaPheArgAlaLeuGluGlnAspLeuProValAsnIleLysPheIleLeuGluGlyMet 198

Qy    568 GATGAGGAGTGCATCAGGGACAGGGGCCTCAGAGGATCTCAGCCCCTGCTACAGCAAGG 624
              |||||:::                :::
Db    199 GluGlu---AlaGlySerValAlaLeuGluGluLeuValGluLysGlyLys 214

Qy    625 ----GGCGTGCCAGTAGCCTTCATTGTGGACGAGGGGGCTTCATCTTG 669
              |||||                    :::|||||||
Db    215 AspArgPhePheSerGlyValAsp------TyrIleVal 227

Qy    670 GATGATTTCATTCCTAAACTTCAAGAAGCCCCATCCGCCTTGAATTGCAGTGTCTCAGAGAAGGGT 729
              |||::::::||                         :::|||               |||
Db    228 ---ThrPheGlyGlyIleLeuHisGluProMetAlaAspLeuLeuLeuLeu 245

Qy    730 TCCATGAAACCTCATGCTGCAAGTAAACATGACTTCAGGC----CACTCTTCAGCTCCT 783
              :::               |||:::|||                           :::
Db    246 AsnSerTyrPheMetValGlyValLysCysArgAspPheHisSerGly---- 263

Qy    784 CCAAAGGAGACAAAGCAATTGGGCATCCTTGCAGCCTGCTGTCAGCCGGATTGGAGCAGACACCA 843
              |||     :::||||||               :::||||||               |||
Db    264 ------ProGlyIleTyrAspGluValVal 280

Qy    844 ATGCCTATCATTATTTGGAAGCGGGACAGTGGTGACTGTTAATTGCACCAACTGGCAAAATGAG 903
              :::|||       |||     |||
Db    281 GlySerLeuValAlaSerProSerGlyHisIleLeuVal 292

Qy    904 TTTCCCTTCCCTGTCAAATATAATCCTGAGCAACCCATGGCCTATTGTGGAACCACTTATAAGC 963
              ::::::|||               :::||||||               |||
Db    293 ------ProGlyLeuTyrAspGluValVal 300

Qy    964 AGGTTTATGGAGAGAAAATCCCTTAACCAATGCAATAATAGGACACCACGGGCACTCACC 1023
              ||||||                    :::|||               |||
Db    301 ------ProLeuThrArgGluGluAsnThr 309

Qy   1024 ATATTCAAAGCAGGGGTCAAGTTCAAATGTCATCCCCCCAGTGGCCCCAGGCCACAGTCAC 1083
              :::||||||               |||
Db    310 ---TyrLysAla 312

Qy   1084 TTCCCGGATTCAC----CCTGGACAGACAGTCCAAGAGGTCCTTA 1122
              :::                    |||:::::||
Db    313 ----IleHisLeuAspLeuGluGlyTyrArgAsnSerArgValGlyLysPheLeu 330

Qy   1123 GAACTCACGAAGAAACAATGTGGGCTGATAACAGAGTCCAGTTCCATGTGTTGAGT------ 1176
              |||||||                           :::     |||||||
Db    331 PheAspThrLysGlyIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis 350

Qy   1177 ------GCCTTTGACCCCCTCCCCGTCAGCCCTTCTGAGACAAGGCCTTGGGC 1224
              |||||||               :::               |||
Db    351 GlyIleLeuGluGlyValAlaPheAsp---------------GluProGlyThrLysThrValIlePro 366

Qy   1225 TACCAGCTGCTCCCGCCGAACCGTACAGTCCGTCTTCCCGGAAGTCAAATATTACTGCC--- 1281
              :::|||               :::||||||               |||
Db    367 GlyArgValIleGlyLysPheSerIleArgLeuValProHisMetAsnValSerArgAlaVal 386

Qy   1282 CCAGTTACTTCGTTCAGTTGGAACACAGACGCCGA 1314
              |||                    :::::||
Db    387 GluLysGlnValThrArgHisLeuGluAspValPheSerLysSerArgAsnSerAsnLys 406
```

Qy   1315 TTCTTTACAAACCTCACCACTGGCATCTAC 1344
           ::::::|||| ||:|:||| :|:|:::::
Db   407 MetValSerMetThrLeuGlyIleuHis 416

RESULT 15
US-10-035-855-57

Sequence 57, Application US/10035855
Publication No. US20030008348A1
GENERAL INFORMATION:
APPLICANT: Desnoyers,Luc
APPLICANT: Eaton,Dan L.
APPLICANT: Goddard,Audrey
APPLICANT: Godowski,Paul J.
APPLICANT: Gurney,Austin L.
APPLICANT: Pan,James
APPLICANT: Stewart,Timothy A.
APPLICANT: Watanabe,Colin K.
APPLICANT: Wood,William I.
APPLICANT: Zhang,Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3030R1C4
CURRENT APPLICATION NUMBER: US/10,035,855
CURRENT FILING DATE: 2001-12-26
PRIOR APPLICATION NUMBER: 60/085579
PRIOR FILING DATE: 1998-05-15
PRIOR APPLICATION NUMBER: 60/112514
PRIOR FILING DATE: 1998-12-15
PRIOR APPLICATION NUMBER: 60/113300
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: 60/113430
PRIOR FILING DATE: 1998-12-23
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-12
PRIOR APPLICATION NUMBER: 60/116843
PRIOR FILING DATE: 1999-01-22
PRIOR APPLICATION NUMBER: 60/125774
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125778
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/125826
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: 60/127035
PRIOR FILING DATE: 1999-03-24
PRIOR APPLICATION NUMBER: 60/127706
PRIOR FILING DATE: 1999-03-31
PRIOR APPLICATION NUMBER: 60/129122
PRIOR FILING DATE: 1999-04-05
PRIOR APPLICATION NUMBER: 60/130359
PRIOR FILING DATE: 1999-04-13
PRIOR APPLICATION NUMBER: 60/131270
PRIOR FILING DATE: 1999-04-21
PRIOR APPLICATION NUMBER: 60/131272
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/131291
PRIOR FILING DATE: 1999-04-27
PRIOR APPLICATION NUMBER: 60/132371
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132379
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/132383
PRIOR FILING DATE: 1999-05-04
PRIOR APPLICATION NUMBER: 60/135750
PRIOR FILING DATE: 1999-05-25
PRIOR APPLICATION NUMBER: 60/138166
PRIOR FILING DATE: 1999-06-08
PRIOR APPLICATION NUMBER: 60/144791
PRIOR FILING DATE: 1999-06-08

PRIOR APPLICATION NUMBER: 60/146970
PRIOR FILING DATE: 1999-07-20
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: 60/162506
PRIOR FILING DATE: 1999-08-03
PRIOR APPLICATION NUMBER: 09/311832
PRIOR FILING DATE: 1999-08-25
PRIOR APPLICATION NUMBER: 09/380142
PRIOR FILING DATE: 1999-10-29
PRIOR APPLICATION NUMBER: 09/644848
PRIOR FILING DATE: 2000-08-22
PRIOR APPLICATION NUMBER: 09/747259
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: 09/816744
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/854208
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/854280
PRIOR FILING DATE: 2001-05-10
PRIOR APPLICATION NUMBER: 09/874503
PRIOR FILING DATE: 2001-06-05
PRIOR APPLICATION NUMBER: 09/865599
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: 09/908,827
PRIOR FILING DATE: 2001-07-18
PRIOR APPLICATION NUMBER: PCT/US99/10733
PRIOR FILING DATE: 1999-05-14
PRIOR APPLICATION NUMBER: PCT/US99/28551
PRIOR FILING DATE: 1999-12-02
PRIOR APPLICATION NUMBER: PCT/US99/30720
PRIOR FILING DATE: 1999-12-22
PRIOR APPLICATION NUMBER: PCT/US00/05601
PRIOR FILING DATE: 2000-03-01
PRIOR APPLICATION NUMBER: PCT/US00/05841
PRIOR FILING DATE: 2000-03-02
PRIOR APPLICATION NUMBER: PCT/US00/14042
PRIOR FILING DATE: 2000-05-22
PRIOR APPLICATION NUMBER: PCT/US00/15264
PRIOR FILING DATE: 2000-06-02
PRIOR APPLICATION NUMBER: PCT/US00/23522
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: PCT/US00/23328
PRIOR FILING DATE: 2000-08-24
PRIOR APPLICATION NUMBER: PCT/US00/32678
PRIOR FILING DATE: 2000-12-01
PRIOR APPLICATION NUMBER: PCT/US00/34956
PRIOR FILING DATE: 2000-12-20
PRIOR APPLICATION NUMBER: PCT/US01/06520
PRIOR FILING DATE: 2001-02-28
PRIOR APPLICATION NUMBER: PCT/US01/17800
PRIOR FILING DATE: 2001-06-01
PRIOR APPLICATION NUMBER: PCT/US01/19692
PRIOR FILING DATE: 2001-06-20
PRIOR APPLICATION NUMBER: PCT/US01/21066
PRIOR FILING DATE: 2001-06-29
PRIOR APPLICATION NUMBER: PCT/US01/21735
PRIOR FILING DATE: 2001-07-09
NUMBER OF SEQ ID NOS: 80
SEQ ID NO 57
   LENGTH: 507
   TYPE: PRT
   ORGANISM: Homo Sapien
US-10-035-855-57

Alignment Scores:
Pred. No.:
Score: 136.50
Percent Similarity: 37.30%
Best Local Similarity: 21.62%
Query Match: 4.94%
DB:

US-10-014-896-1 (1-1509) x US-10-035-855-57 (1-507)

Length: 507
Matches: 80
Conservative: 58
Mismatches: 127
Indels: 105
Gaps: 17

0.00165

```
Qy    334 GGCTCGGACCCCAGCTTGCAGCCTACCTCCTGAITGGTGGCTCACTTTGATGTGGTCCTGCC 393
          ||||||||||||||||||||||||||    :::
Db    119 GlySerAspProThrLysGlyThrValCysPheTyrGlyHisLeuAspValGlnProAla 138

Qy    394 CCTGAA---GAAGGCTGGGAGGTGCCCCAITTCTCTGGGTTGGAGCGTGATGGGGTCAIC 450
          :::||||||     |||||||| |||||  :::
Db    139 AspArgGlyAspGlyTrpLeuThrAspProTyrValLeuThrGluValAspGlyLysLeu 158

Qy    451 TATGGTPCGGGGGCCACACTGGCACGACAAAGAACTCTGTGATGGCATTACTGCGGCCTTGGAG 510
          ||||||| |||||     |||    |||   |||||
Db    159 TyrGlyArgGlyAlaThrAspAspAsnLysGlyValLeuAlaIleLeuAsnAlaValSer 178

Qy    511 CTCCTG---CTGAICAGGAAGTACAAICCCCGGAAGAICTTTCTTCAITTCTCTGGGCCAIT 567
          :::     :::|||     ::: |||||    |||
Db    179 AlaPheArgAlaLeuGlnAspLeuProValAsnIleLysPhelleIleGluGlyMet 198

Qy    568 GATGAGGAGTCAICAGGGACAGGGGGCTCAGGAGGAICTCAGCCCTGCTACAGTCAAGG--- 624
          ::: |||||     |||
Db    199 GluGlu-------AlaGlySerValAlaLeuGluGluLeuValGluLysGlyLys 214

Qy    625 -------GGCGTCCCAGCGTAGCCTTCAITTGTGACGAGGGGGCTTCAITCTTG 669
          :::||||||| |||
Db    215 AspArgPhePheSerGlyValAsp------TyrIleVal 227

Qy    670 GAIGATTTCAITCCTAACTTCAAGAGAGCCCAICGGCCTTGAITGTGCAGTCTCAGAGAAGGGT 729
          ||| :::||||||    ::: |||||    :::||| |||
Db    228 AspAsnLeuTrpIleSerGlnArgGlyLysPro-----AlaIleThrThrGlyThrArgGly 245

Qy    730 TCCAITGAAACCTCAICGCTGCAAGTAAACAITGATCAIGAICTTCAGGC-----CACTCITCAGCTCCT 783
          :::|||||||||    ::: :::|||
Db    246 AsnSerTyrPheMetValGlyValLysCysArgAspGlnAspPhelleSerGly------ 263

Qy    784 CCAAAGGACAICAGCAITGGCAITCCITGCAICTGCTGITCAICAGCCGAITGGAGCAGACACCA 843
          |||     ::: :::||||||     :::
Db    264 -------ThrPheGlyIleLeuHisGluProMetAlaAspLeuValAlaLeuLeu 280

Qy    844 AITGCCTAICAITAITTGGAAGCGGGACAGTGGTGACTGITAITGCAICGAACTGCAAAITGAG 903
          |||||||||||||
Db    281 GlySerLeuValAspSerSerGlyHisIleIleLeuVal------ 292

Qy    904 TITCCCTICCCTGTCAATAITAAITCCTGAGCAACCCAITGGGTAITTGAACCACTTAITAAGC 963
          |||     :::::::
Db    293 -------ProGlyIleTyrAspGluValVal 300

Qy    964 AGGTTTAITGGAGAGAAAITCCITAACCAAITGCAAITAITCAGGACCACCAGGGCAICTCACC 1023
          ||||||||||     |||
Db    301 -------ProLeuThrGluGlyIleuIleAsnThr--- 309

Qy   1024 AITAITTCAAAGCAGGGGTCAIGTCAAAITGCAITCCCCCAGTGGCCCAGGGCCACAIGTCAAC 1083
          :::|||||
Db    310 ---TyrLysAla------ 312

Qy   1084 TICCGGAITCAIC---- 1122
          |||||||
Db    313 -----TleHisLeuAspLeuGluGluGluTyrArgValGlyLysPheLeu 330

Qy   1123 GAACTCACGAIAGGAICAITTGTGGCTGAITAACAIGAGTCCAIGTGITTGAIGT----- 1176
          |||||     :::
Db    331 PheAspThrLysGluGluIleLeuMetHisLeuTrpArgTyrProSerLeuSerIleHis 350

Qy   1177 -----GCCTTTGACCCCGTCCCCGTCCAIGCCCTTCTCGAITGACAIGAGGCCTTGGGC 1224
          |||||||||||||    :::
Db    351 GlyIleGluGlyAlaPheAsp-------GluProGlyThrLysThrValIlePro 366

Qy   1225 TACCAIGCTGCTCCGACAICCGTCTTCCCGGAIGTCAITAITTAICTGCC--- 1281
          :::::::||||| |||||||||||||
Db    367 GlyArgValIleGlyLysPheSerIleLeuArgLeuValProHisMetAsnValSerAlaVal 386

Qy   1282 ------CCAGTTACTTCTTAITTGGGAICAICACAICAGAICAGCCGA 1314
          |||     ::::::
Db    387 GluLysGlnValThrArgHisLeuGluGluAspValPheSerLysArgAsnSerSerAsnLys 406
```

```
Qy   1315 TTCTTTACAAACCTCACCACTGGCATCTAC 1344
          :::::::|||     :::::::
Db    407 MetValSerMetThrLeuGlyLeuHis 416
```

Search completed: June 27, 2003, 11:26:24
Job time : 92 secs